PERBANDINGAN METODE ESTIMASI PARAMETER *MAXIMUM LIKELIHOOD* DAN BAYESIAN PADA REGRESI *RIDGE*
(Studi Kasus Rata-rata Harapan Sekolah Kota/ Kabupaten di Jawa Timur Tahun 2015 dan Kepadatan Penduduk menurut Kecamatan di Kabupaten Tulungagung Tahun 2016)

SKRIPSI

oleh:
DIAH AYU WIDYASTUTI
145090507111023

PROGRAM STUDI SARJANA STATISTIKA
JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2018

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

repository.ub.ac.id

PERBANDINGAN METODE ESTIMASI PARAMETER *MAXIMUM LIKELIHOOD* DAN BAYESIAN PADA REGRESI *RIDGE*

(Studi Kasus Rata-rata Harapan Sekolah Kota/ Kabupaten di Jawa Timur Tahun 2015 dan Kepadatan Penduduk menurut Kecamatan di Kabupaten Tulungagung Tahun 2016)

SKRIPSI

Sebagai salah satu syarat untuk memperoleh gelar Sarjana Statistika

oleh:
DIAH AYU WIDYASTUTI
145090507111023

PROGRAM STUDI SARJANA STATISTIKA
JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2018

UNIVERSITAS BRAWIJAYA

repository.ub.ac.id

LEMBAR PENGESAHAN SKRIPSI

PERBANDINGAN METODE ESTIMASI PARAMETER *MAXIMUM LIKELIHOOD* DAN BAYESIAN PADA REGRESI *RIDGE*
(Studi Kasus Rata-rata Harapan Sekolah Kota/ Kabupaten di Jawa Timur Tahun 2015 dan Kepadatan Penduduk menurut Kecamatan di Kabupaten Tulungagung Tahun 2016)

oleh:
DIAH AYU WIDYASTUTI
145090507111023

Setelah dipertahankan di depan Majelis Penguji
Pada tanggal 13 Juli 2018
Dan dinyatakan memenuhi syarat untuk memperoleh gelar
Sarjana Statistika

Dosen Pembimbing

Achmad Efendi, S.Si., M.Sc., Ph.D
NIP. 198102192005011001

Mengetahui,
Ketua Jurusan Statistika
Fakultas MIPA
Universitas Brawijaya

Rahma Fitriani, S.Si., M.Sc., Ph.D
NIP. 197603281999032001

UNIVERSITAS BRAWIJAYA

repository.ub.ac.id

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : **Diah Ayu Widyastuti**
NIM : **145090507111023**
Jurusan : **Statistika**
Program Studi : **Sarjana Statistika**
Skripsi berjudul :

PERBANDINGAN METODE ESTIMASI PARAMETER *MAXIMUM LIKELIHOOD* DAN BAYESIAN PADA REGRESI *RIDGE*
(Studi Kasus Rata-rata Harapan Sekolah Kota/ Kabupaten di Jawa Timur Tahun 2015 dan Kepadatan Penduduk menurut Kecamatan di Kabupaten Tulungagung Tahun 2016)

Dengan ini menyatakan bahwa:

1. **Isi dari skripsi yang saya buat adalah benar-benar karya sendiri dan tidak menjiplak karya orang lain, selain nama-nama yang termaktub di isi dan tertulis di daftar pustaka dalam skripsi ini.**
2. **Apabila di kemudian hari ternyata skripsi yang saya tulis terbukti hasil jiplakan, maka saya akan bersedia menanggung segala resiko yang akan saya terima.**

Demikian pernyataan ini saya buat dengan segala kesadaran.

Malang, 24 Juli 2018
Yang menyatakan,

Diah Ayu Widyastuti
145090501111025

UNIVERSITAS BRAWIJAYA

PERBANDINGAN METODE ESTIMASI PARAMETER *MAXIMUM LIKELIHOOD* DAN BAYESIAN PADA REGRESI *RIDGE*

(Studi Kasus Rata-rata Harapan Sekolah Kota/ Kabupaten di Jawa Timur Tahun 2015 dan Kepadatan Penduduk menurut Kecamatan di Kabupaten Tulungagung Tahun 2016)

ABSTRAK

Asumsi non multikolinieritas dalam regresi linier berganda diperlukan agar menghasilkan estimasi parameter yang akurat dan dapat diinterpretasi dengan baik. Apabila antara variabel prediktor terdapat korelasi yang tinggi menunjukkan bahwa asumsi tersebut tidak terpenuhi. Salah satu metode yang dapat digunakan untuk menangani multikolinieritas adalah regresi *ridge*, di mana metode estimasi yang digunakan yaitu *Maximum Likelihood Estimation* (MLE) dan Bayesian dengan tujuan mengurangi tingkat multikolinieritas dan menghasilkan estimasi parameter yang lebih akurat. Data penelitian menggunakan dua data sekunder yaitu data Rata-rata Harapan Sekolah pada Kabupaten/ Kota di Jawa Timur tahun 2015 dan Kepadatan Penduduk menurut Kecamatan di Kabupaten Tulungagung tahun 2016. Hasil estimasi parameter dengan MLE terbukti dapat menangani multikolinieritas, ditunjukkan dari nilai *Variance Inflation Factor* (VIF) akhir yang kurang dari 10. Jika dilakukan perbandingan antara MLE dan Bayesian berdasarkan kriteria kebaikan model maka disimpulkan bahwa metode Bayesian lebih baik dalam estimasi parameter pada regresi *ridge*. Kriteria kebaikan model menunjukkan nilai R^2_{adj} metode Bayesian lebih besar daripada MLE yaitu untuk Data 1 sebesar 70.60% sedangkan pada Data 2 sebesar 93.70%, selain itu nilai MSE (*Mean Square Error*), BIC (*Bayesian Information Criterion*) dan AIC (*Akaike Information Criterion*) menunjukkan nilai yang lebih kecil pada metode Bayesian.

Kata kunci: Bayesian, MLE, Multikolinieritas, Regresi *Ridge*, VIF

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

COMPARISON OF MAXIMUM LIKELIHOOD AND BAYESIAN PARAMETER ESTIMATION METHOD ON RIDGE REGRESSION
(Case Study Average School Life Expectancy of City/ Regency in East Java in 2015 and Citizen Density in Each District of Tulungagung in 2016)

ABSTRACT

The assumption of non multicollinierity in multiple linier regression is required to produce good interpretation of parameters. A high correlation between predictors indicates that the assumption is violated. One of methods that can handle multicollinierity is ridge regression, where the estimation method used is Maximum Likelihood Estimation (MLE) and Bayesian. In this research, we use two secondary data, they are Average School Life Expectancy City/ Regency in East Java in 2015 and Citizen Density in Each District of Tulungagung in 2016. Result of parameter estimation shows that MLE can handle multicollinierity with Variance Inflation Factor (VIF) less than 10. Comparison between MLE and Bayesian based on goodness of fit model shows that Bayesian is better than MLE for parameter estimation on ridge regression. It can be shown from goodness of fit model, the value of R^2_{adj} is bigger than MLE method for Data 1, 70.60% and Data 2, 93.70%. Additionally, MSE (Mean Square Error), BIC (Bayesian Information Criterion), and AIC (Akaike Information Criterion) show smaller values for Bayesian method.

Keywords: Bayesian, MLE, Multicollinearity, Ridge Regression, VIF

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

KATA PENGANTAR

Puji syukur kepada Allah SWT atas rahmat dan hidayah-Nya sehingga skripsi yang berjudul **Perbandingan Metode Estimasi Parameter *Maximum Likelihood* dan Bayesian pada Regresi *Ridge*** dapat terselesaikan dengan baik. Dalam penulisan ini, tidak terlepas dari bantuan dan dukungan dari berbagai pihak, untuk itu penulis menyampaikan rasa terima kasih kepada:

1. Bapak Achmad Efendi, S.Si., M.Sc., Ph.D selaku dosen pembimbing skripsi atas waktu, saran, dan bimbingan yang telah diberikan.
2. Ibu Ir. Heni Kusdarwati, MS selaku anggota majelis penguji I atas waktu, saran, dan bimbingan yang telah diberikan.
3. Ibu Dr. Suci Astutik, S.Si., M.Si selaku anggota majelis penguji II atas waktu, saran, dan bimbingan yang telah diberikan.
4. Ibu Rahma Fitriani, S.Si., M.Sc., Ph.D selaku Ketua Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam.
5. Bapak, Mamah, Kakak beserta seluruh keluarga besar yang selalu memberikan dukungan, doa, dan motivasi.
6. Teman-teman terdekat Imron, Ayu, Asna, Lina, Hazrina, Zahra, AG squad (Rinintya, Ivana, Silvia, dan Rike) serta seluruh teman-teman statistika 2014 yang senantiasa memberikan dukungan, doa, semangat, dan motivasi.
7. Semua pihak yang telah membantu baik secara langsung dan tidak langsung dalam penyelesaian skripsi.

Penulis menyadari bahwa skripsi ini masih kurang sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun untuk penulisan yang lebih baik. Penulis berharap semoga skripsi ini bermanfaat bagi seluruh pihak yang membaca.

Malang, Juli 2018

Penulis

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

DAFTAR ISI

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

DAFTAR GAMBAR

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

DAFTAR TABEL

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1. Latar Belakang

Salah satu cara yang dapat digunakan untuk mengetahui hubungan antara variabel yang terlibat dalam suatu penelitian adalah analisis regresi. Variabel-variabel yang terlibat meliputi variabel prediktor dan variabel respon, di mana variabel prediktor menjadi faktor yang mempengaruhi variabel respon. Analisis regresi yang melibatkan beberapa variabel prediktor (X_1, X_2, X_3, ..., X_p) dan satu variabel respon (Y) adalah analisis regresi linier berganda. Penerapan analisis regresi dapat digunakan untuk mengetahui hubungan antara variabel prediktor dan variabel respon yang bersifat positif maupun negatif serta dapat digunakan untuk melakukan estimasi terhadap nilai dari variabel respon apabila terjadi kenaikan maupun penurunan pada variabel prediktor.

Terdapat beberapa asumsi yang harus terpenuhi dalam regresi linier berganda, yaitu kenormalan galat, kehomogenan ragam galat, non autokorelasi dan non multikolinieritas. Non multikolinieritas menjadi salah satu asumsi yang sering tidak terpenuhi dalam suatu analisis, hal ini dikarenakan variabel prediktor yang terlibat dalam suatu penelitian memiliki korelasi yang tinggidengan variabel prediktor lainnya. Korelasi antara beberapa variabel prediktor ini menyebabkan estimasi parameter bersifat bias dan nilai ragam menjadi besar.

Menurut Kurtner *et al* (2004) terdapat beberapa metode dalam estimasi parameter pada regresi linier yaitu dengan menggunakan Metode Kuadrat Terkecil (MKT) dan *Maximum Likelihood Estimation* (MLE). MKT bekerja dengan cara meminimumkan Jumlah Kuadrat Galat (JKG), sedangkan MLE memaksimumkan fungsi peluang dari distribusi data sampel. Kedua metode ini memiliki hasil yang sama ketika diterapkan untuk mengestimasi parameter pada regresi linier sederhana maupun linier berganda. Ketika variabel prediktor memiliki korelasi yang tinggi dengan variabel prediktor yang lain maka MKT dan MLE kurang tepat untuk memperoleh hasil estimasi yang akurat, karena menyebabkan model yang terbentuk dari analisis regresi tidak lagi efisien. Hal tersebut dapat dilihat dari nilai *standard error* koefisien regresi menjadi sangat besar (*overestimate*) dengan kata lain mengurangi tingkat presisi dari estimasi, adanya multikolinieritas memungkinkan menghasilkan estimasi koefisien regresi dengan tanda yang salah (di mana yang seharusnya positif bisa negatif dan sebaliknya).

Akibat dari multikolinieritas menyebabkan kurang tepat hasil estimasi menggunakan MKT dan MLE, sehingga masalah multikolinieritas dapat diatasi dengan beberapa cara, antara lain menghilangkan satu atau beberapa variabel yang memiliki tingkat korelasi yang sangat tinggi, menambah banyak observasi untuk mengurangi korelasi antara variabel prediktor, selain itu juga dapat digunakan metode regresi *ridge*. Metode regresi *ridge* merupakan suatu metode yang memodifikasi matriks $\boldsymbol{X'X}$ dengan menambah konstanta bias (k) ke dalam diagonal matriks $\boldsymbol{X'X}$ sehingga estimasi parameter yang dihasilkan dapat menurunkan nilai *standard error* hasil estimasi. Menurut Hestie *et al* (2008) metode regresi *ridge* digunakan apabila variabel prediktor yang terlibat dalam suatu penelitian memiliki korelasi antara variabel prediktor yang satu dan variabel prediktor yang lain. Regresi *ridge* pertama kali diperkenalkan oleh Hoerl dan Kennard pada tahun 1970, metode ini digunakan untuk menangani kasus multikolinieritas pada data hasil pengamatan di mana terdapat korelasi yang tinggi antara variabel prediktor.

Metode estimasi parameter pada regresi *ridge* dapat menggunakan MKT dan MLE yang telah dimodifikasi dengan penambahan tetapan bias (k). Penggunaan tetapan bias (k) dapat dilakukan dengan beberapa cara yang diperkenalkan oleh Hoerl and Kennard (1970); Hoerl, Kennard and Balwin (1975); Lawless and Wang (1976), dsb. Selain MKT dan MLE, metode Bayesian dengan metode simulasi *Markov Chain Monte Carlo* (MCMC) juga dapat digunakan untuk pendugaan parameter pada regresi *ridge*. Penggunaan metode Bayesian cenderung lebih efisien daripada MKT dan MLE karena pada metode Bayesian mempertimbangkan informasi distribusi data sampel dan memperhitungkan informasi awal yang disebut distribusi prior, sedangkan pada MKT dan MLE dalam estimasi parameter hanya memanfaatkan informasi dari data sampel.

Penelitian terdahulu Abidin (2015) metode Bayesian digunakan pada estimasi parameter regresi *ridge* sebagai langkah penanganan multikolinieritas dan menunjukkan bahwa metode Bayesian lebih efisien dibandingkan dengan MKT yang ditunjukkan dari nilai

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

standard error dari metode Bayesian lebih kecil. Penelitian Effrihan (2017) membahas estimasi regresi *ridge* dengan metode Bayesian yang memanfaatkan data bangkitan untuk mengetahui kriteria penduga dan menunjukkan bahwa metode Bayesian sangat baik digunakan dengan ukuran sampel yang kecil dilihat dari nilai *standard error*, MSE yang lebih kecil dari pada metode MKT, namun untuk data dengan observasi yang lebih besar hasil MKT dan Bayesian cenderung sama dalam estimasi regresi *ridge*. Penelitian ini membahas mengenai estimasi parameter regresi *ridge* sebagai langkah untuk mengatasi masalah multikolinieritas dengan MLE dan metode Bayesian pada dua data yang mengandung multikolinieritas. Dalam penelitian ini terdapat asumsi regresi yang tidak terpenuhi sehingga metode Bayesian bisa digunakan, karena pada metode Bayesian sendiri tidak mempertimbangkan asumsi dalam estimasi parameter. Dua data yang mengandung multikolinieritas berasal dari pengamatan yang memiliki jumlah amatan yang berbeda, sehingga dapat diketahui bagaimana hasil estimasi parameter dari MLE dan Bayesian pada regresi *ridge*.

1.2. Rumusan Masalah

Berdasarkan latar belakang, permasalahan yang dapat dirumuskan pada penelitian ini adalah:

1. Bagaimana hasil penerapan metode estimasi parameter dengan *Maximum Likelihood* dan Bayesian pada regresi *ridge*?
2. Bagaimana perbandingan hasil estimasi parameter antara *Maximum Likelihood* dan Bayesian pada regresi *ridge*?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan pada sub bab sebelumnya, maka tujuan dari penelitian ini adalah:

1. Menerapkan metode estimasi parameter dengan *Maximum Likelihood* dan Bayesian pada regresi *ridge*.
2. Membandingkan hasil estimasi parameter antara *Maximum Likelihood* dan Bayesian pada regresi *ridge* menggunakan R^2_{adj}, *Mean Square Error* (MSE), *Akaike Information Criterion* (AIC), dan *Bayesian Information Criterion* (BIC).

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

1.4. Batasan Masalah

Batasan masalah yang ditetapkan pada penelitian ini adalah:

1. Penelitian bertujuan untuk menangani masalah asumsi non multikolinieritas yang tidak terpenuhi.
2. Menggunakan distribusi data normal sebagai penyusun distribusi prior dan posterior.

1.5. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah dapat memberikan informasi mengenai penerapan metode estimasi parameter dengan *Maximum Likelihood* dan Bayesian pada regresi *ridge.*

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

BAB II
TINJAUAN PUSTAKA

2.1. Analisis Regresi Linier Berganda

Analisis regresi linier terbagi menjadi dua, yaitu analisis regresi linier sederhana dan analisis regresi linier berganda. Analisis regresi linier sederhana adalah analisis yang melibatkan satu variabel prediktor dan satu variabel respon, sedangkan analisis regresi linier berganda adalah analisis regresi yang melibatkan beberapa variabel prediktor dan satu variabel respon. Analisis regresi linier memiliki beberapa asumsi yang harus dipenuhi yaitu kenormalan galat, kehomogenan ragam galat, non autokorelasi dan non multikolinieritas (Wathen *et al,* 2014).

Model dikatakan linier apabila parameter bersifat linier atau pangkat tertinggi dari parameter adalah satu (Draper dan Smith, 1992). Hubungan antara variabel-variabel dalam analisis regresi linier dapat dirumuskan dalam sebuah persamaan. Menurut Pramoedyo (2013) rumusan hubungan variabel yang terlibat dalam analisis regresi linier dapat dilihat pada persamaan (2.1):

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \varepsilon_i \qquad (2.1)$$

di mana:

y_i	:	variabel respon pada nilai pengamatan ke-i
β_0	:	titik perpotongan antara suatu garis regresi dengan sumbu y (intersep)
$\beta_1, \ldots, \beta_p$	:	koefisien regresi untuk setiap variabel prediktor ke-p
X_{ip}	:	nilai pengamatan ke-i pada variabel prediktor ke-p
ε_i	:	galat pada pengamatan ke-i
n	:	banyak pengamatan
p	:	banyak variabel prediktor
i	:	1, 2, ..., n
j	:	1, 2, ..., p

2.2. Asumsi Klasik Regresi Linier Berganda

2.2.1. Normalitas Galat

Salah satu asumsi yang harus terpenuhi dalam analisis regresi linier adalah asumsi normalitas galat, di mana galat (ε_i) yang merupakan variabel acak yang berdistribusi normal dengan nilai tengah 0 dan ragam σ^2.

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

Hipotesis yang digunakan dalam pengujian asumsi normalitas galat adalah:

H_0: $\varepsilon \sim N\ (\mu, \sigma^2)$ *vs*

H_1: $\varepsilon \nsim N\ (\mu, \sigma^2)$

Pengujian asumsi normalitas galat salah satunya dapat menggunakan statistik uji *Kolmogorov Smirnov* (Pramoedyo, 2013).

$$D = \text{maksimum}\left|F_0\left(X_i\right) - S_n\left(X_i\right)\right|; i = 1, 2, \ldots, n \qquad (2.2)$$

Kriteria pengambilan keputusan terima H_0 apabila ($D < D_\alpha$) yang berarti nilai statistik uji D kurang dari nilai D pada titik kritis dengan nilai taraf nyata (α) tertentu, sehingga dapat diambil keputusan bahwa galat berdistribusi normal.

2.2.2. Homogenitas Ragam Galat

Asumsi kehomogenan ragam galat menunjukkan bahwa seiring meningkatnya rata-rata maka ragam seharusnya tetap konstan, tetapi terdapat kemungkinan semakin naik nilai rata-rata menyebabkan nilai ragam juga naik, sehingga perlu dilakukan pengujian asumsi kehomogenan ragam galat. Pengujian asumsi dilakukan agar hasil penduga yang diperoleh efisien.

Menurut Gujarati (2010) pengujian kehomogenan ragam galat bisa dilakukan dengan statistik uji *Breusch Pagan* dengan hipotesis sebagai berikut:

H_0: $\alpha_0 = \alpha_1 = \ldots = \alpha_p$ *vs*

H_1: $\alpha_0 \neq \alpha_1 \neq \ldots \neq \alpha_p$

Statistik uji *Breusch Pagan* dapat dilihat pada persamaan (2.3):

$$LM = nR^2 \qquad (2.3)$$

Kriteria pengambilan keputusan terima H_0 apabila nilai statistik uji LM < $\chi^2_{(\alpha/2,p)}$ yang berarti bahwa asumsi kehomogenan ragam galat terpenuhi.

2.2.3. Non Autokorelasi

Pengujian asumsi non autokorelasi bertujuan untuk mengetahui pada beberapa pengamatan memiliki galat yang saling berkorelasi atau tidak. Jika terjadi kovarian dan korelasi antar galat tidak sama dengan nol, maka hal ini dapat dikatakan sebagai pelanggaran asumsi. Metode pengujian asumsi non-autokorelasi dapat dilakukan dengan metode *Durbin Watson* dengan hipotesis sebagai berikut:

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

$H_0: \rho = 0$ *vs*
$H_1: \rho \neq 0$

Statistik uji yang digunakan dalam metode *Durbin Watson* sesuai dengan persamaan (2.4):

$$d = \frac{\sum_{i=2}^{n} \left(\hat{\varepsilon}_i - \hat{\varepsilon}_{i-1} \right)^2}{\sum_{i=1}^{n} \hat{\varepsilon}_i^{\,2}} \tag{2.4}$$

di mana:
$\hat{\varepsilon}_i$: nilai galat pada pengamatan ke-i
$\hat{\varepsilon}_{i-1}$: nilai galat pada pengamatan sebelumnya yaitu ke-(i-1)

Kriteria dalam pengambilan keputusan pada statistik uji *Durbin Watson* dapat dilihat pada Tabel 2.1:

Tabel 2.1 Kriteria Pengambilan Keputusan Statistik Uji *Durbin Watson*

No.	Hipotesis Nol (H_0)	Keputusan	Nilai Statistik Uji
1.	Tidak terdapat autokorelasi positif	Tolak H_0	$0 < d < d_l$
2.	Tidak terdapat autokorelasi positif	Ragu-ragu	$d_l < d < d_u$
3.	Tidak terdapat autokorelasi positif	Tolak H_0	$4\text{-}d_l < d < 4$
4.	Tidak terdapat autokorelasi positif	Ragu-ragu	$4\text{-}d_u \leq d \leq 4\text{-}d_l$
5.	Tidak terdapat autokorelasi positif	Terima H_0	$d_u < d < 4\text{-}d_u$

2.2.4. Non Multikolinieritas

Salah satu asumsi yang harus terpenuhi dalam analisis regresi linier yaitu asumsi non multikolinieritas, di mana asumsi non multikolinieritas menyatakan bahwa antara variabel prediktor dalam suatu model yang dihasilkan harus bersifat saling bebas satu sama lain. Variabel prediktor yang berhubungan atau berkorelasi antara satu dengan yang lain menyebabkan kesulitan untuk mengambil kesimpulan mengenai masing-masing koefisien regresi beserta pengaruhnya terhadap variabel respon, selain itu hubungan antara variabel prediktor menyebabkan ragam menjadi besar meskipun tidak bias.

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

Untuk mengetahui antar variabel prediktor memiliki korelasi atau tidak, maka salah satu cara mendeteksinya dapat dilakukan dengan melihat nilai *Variance Inflation Factor* (VIF). VIF merupakan ukuran keragaman total dari salah satu variabel yang dapat dijelaskan oleh keragaman variabel prediktor lain. VIF dapat dihitung dengan persamaan (2.5):

$$\mathrm{VIF}_j = \frac{1}{1-R_j^2};\ j = 1, 2, \ldots, p \qquad (2.5)$$

di mana:

R_j^2: koefisien determinasi dari regresi antara variabel prediktor ke-j dengan semua variabel prediktor yang lain.

Koefisien determinasi diperoleh dengan dari hasil metode *auxiliary regression* yaitu dengan meregresikan masing-masing variabel prediktor dengan variabel prediktor yang lain. Prinsip dari penggunaan metode *auxiliary regression* adalah salah satu variabel prediktor berperan sebagai variabel respon dan variabel lainnya berperan sebagai variabel prediktor, kondisi ini diulang sampai semua variabel prediktor yang terlibat menjadi variabel respon secara bergantian. Rumus koefisien determinasi yang dihasilkan dapat dilihat pada persamaan (2.6):

$$R_j^2 = \frac{JKR}{JKT} = 1 - \frac{JKG}{JKT} \qquad (2.6)$$

Jika nilai VIF lebih dari 10 sudah cukup bukti untuk menyatakan bahwa terjadi multikolinieritas pada data yang akan diteliti (Wathen *et al*, 2014).

2.3. Estimasi Parameter Regresi Linier Berganda

Maximum Likelihood Estimation (MLE) dapat digunakan sebagai metode untuk mengestimasi nilai koefisien regresi pada regresi linier berganda (Pramoedyo, 2013). Persamaan (2.7) menunjukkan persamaan model estimasi parameter:

$$\hat{y}_i = \hat{\beta}_0 + \sum_{i=1}^{n} \hat{\beta}_j X_{ij} \qquad (2.7)$$

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

Menurut Kurtner *et al* (2004) persamaan (2.1) dapat dirumuskan dalam bentuk persamaan matriks sebagai berikut:

$$\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.8}$$

di mana:

X : matriks yang berisi angka 1 digabung dengan p variabel prediktor sebagai kolom dengan n buah observasi sebagai baris berdimensi $\boldsymbol{n} \times (\boldsymbol{p}+\mathbf{1})$

y : variabel respon yang dibentuk dalam vektor kolom dengan n buah observasi berdimensi $(\boldsymbol{n} \times \mathbf{1})$

β : vektor koefisien regresi sebanyak $p+1$ variabel berdimensi $(\boldsymbol{p}+\mathbf{1}) \times \mathbf{1}$

ε : vektor galat berdimensi $(\boldsymbol{n} \times \mathbf{1})$

Persamaan (2.8) dapat dijabarkan dalam bentuk matriks sebagai berikut:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ 1 & X_{31} & X_{32} & \cdots & X_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2.9}$$

Estimasi bagi koefisien regresi dengan MLE diperoleh dengan memaksimumkan peluang distribusi data sampel, dalam persamaan matriks dinyatakan sebagai berikut:

$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \tag{2.10}$$

Prinsip dasar dalam pembentukan MLE yaitu dengan memaksimumkan fungsi peluang dari distribusi data sampel. Persamaan fungsi *likelihood* dinyatakan sebagai berikut:

$$L(\theta) = L(\theta_1, \theta_2, \ldots, \theta_P) = f(y, \theta) \tag{2.11}$$

Jika semua peluang sampel diambil, maka θ merupakan fungsi dari y dan merupakan variabel acak itu sendiri, berdasarkan penjelasan tersebut dapat disimpulkan bahwa hal ini merupakan estimasi dengan MLE dari θ. Berikut merupakan bentuk perkalian fungsi *likelihood* :

$$L(\theta) = f(y_1|\theta)\, f(y_2|\theta) \ldots . f(y_n\,|\,\theta)$$

$$L(\theta) = \prod_{i=1}^{n} f(y_i\,|\,\theta)$$

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

Pembentukan fungsi *likelihood* pada regresi linier berganda bergantung pada distribusi data sampel. Variabel acak y_i dengan distribusi data sampel normal dimana rata-rata (μ) dan ragam (σ^2) maka fungsi kepekatan peluang distribusi data sampel dapat ditulis sebagai berikut:

$$y_i \sim N\left(\mu, \sigma^2\right)$$

$$f\left(y_i\right) = (2\pi\sigma^2)^{-\frac{1}{2}} exp\left[\frac{\left(y_i - \mu\right)^2}{2\sigma^2}\right]$$

$$= \frac{1}{\sqrt{2\mu\sigma^2}} exp\left[\frac{\left(y_i - \mu\right)^2}{2\sigma^2}\right]$$

Asumsi normalitas menyatakan bahwa (ε_i) memiliki distribusi normal dengan rata-rata (0) dan ragam (σ^2), sehingga $\varepsilon_i \sim N(0, \sigma^2)$ maka fungsi kepekatan peluang bagi e_i adalah sebagai berikut:

$$f(\varepsilon_i) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{\left(\varepsilon_i - 0\right)^2}{2\sigma^2}\right]$$

$$f(\varepsilon_i) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{\left(\varepsilon_i\right)^2}{2\sigma^2}\right]$$

Berikut fungsi kepekatan peluang untuk pembentukan fungsi *likelihood* pada regresi linier berganda sebagai berikut:

$$f\left(y_i | x_i, \beta, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{p}(y_i - \beta_0 - \beta_1 X_{i1} - \ldots - \beta_p X_{ip})^2\right]$$

$$f\left(y_i | x_i, \beta, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2\sigma^2}\left(\mathbf{y} - \mathbf{X\beta}\right)'\left(\mathbf{y} - \mathbf{X\beta}\right)\right]$$

$$f\left(y_i | x_i, \beta, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-1/2} exp\left[-\frac{1}{2\sigma^2}\left(\mathbf{y} - \mathbf{X\beta}\right)'\left(\mathbf{y} - \mathbf{X\beta}\right)\right] \quad (2.12)$$

Menurut Kurtner *et al* (2004) penentuan fungsi *likelihood* pada regresi linier berganda adalah sebagai berikut:

$$L\left(y_i | x_i, \beta, \sigma^2\right) = \prod_{i=1}^{n} f\left(y_i | x_i, \beta, \sigma^2\right)$$

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

$$L\left(y_i|x_i,\beta,\sigma^2\right)=\left(2\pi\sigma^2\right)^{-\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X\beta}\right)'\left(\mathbf{y}-\mathbf{X\beta}\right)\right] \quad (2.13)$$

Fungsi *likelihood* pada persamaan (2.13) selanjutnya diubah dalam bentuk *ln* fungsi *likelihood* sebagai berikut:

$$ln\,L\left(y_i|x_i,\beta,\sigma^2\right)=ln\left(\left(2\pi\sigma^2\right)^{-\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X\beta}\right)'\left(\mathbf{y}-\mathbf{X\beta}\right)\right]\right)$$

$$ln\,L\left(y_i|x_i,\beta,\sigma^2\right)=ln\left(\left(2\pi\sigma^2\right)^{-\frac{n}{2}}\right)+ln\left(\exp\left[-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X\beta}\right)'\left(\mathbf{y}-\mathbf{X\beta}\right)\right]\right)$$

$$ln\ L(y_i\,|\,x_i,\beta,\sigma^2)=-\frac{n}{2}ln(2\pi)\ -\frac{n}{2}ln(\sigma^2)-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X\beta}\right)'\left(\mathbf{y}-\mathbf{X\beta}\right)$$

$$ln\ L\left(y_i|x_i,\beta,\sigma^2\right)=-\frac{n}{2}ln\left(2\pi\right)\ -\frac{n}{2}ln\left(\sigma^2\right)-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X\beta}\right)'\left(\mathbf{y}-\mathbf{X\beta}\right) \quad (2.14)$$

Persamaan (2.14) digunakan untuk membentuk persamaan normal di mana turunan pertama terhadap β sama dengan nol, sehingga diperoleh hasil estimasi parameter dengan MLE untuk regresi linier berganda.

$$\begin{aligned}
\frac{\partial\, ln\,L\left(y_i|x_i,\beta,\sigma^2\right)}{\partial\boldsymbol{\beta}} &= 0\\
-\frac{1}{2\sigma^2}\left(-2\mathbf{X}'\mathbf{Y}+2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}\right) &= 0\\
-2\mathbf{X}'\mathbf{Y}+2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} &= 0\\
-\mathbf{X}'\mathbf{Y}+\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} &= 0\\
\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}'\mathbf{Y}
\end{aligned}$$

$$\boldsymbol{\beta}=\left(\mathbf{X}'\mathbf{X}\right)^{\mathbf{-1}}\mathbf{X}'\mathbf{Y} \quad (2.15)$$

2.4. Multikolinieritas

Multikolinieritas terjadi apabila sebagian variabel prediktor maupun seluruh variabel prediktor pada suatu penelitian mengalami korelasi atau terdapat hubungan linier.

2.4.1. Akibat Multikolinieritas

Tidak terpenuhinya asumsi non multikolinieritas menyebabkan hasil estimasi kurang akurat. Menurut Kurtner *et al* (2004) berikut

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

beberapa akibat dari pelanggaran asumsi multtikolinieritas pada suatu penelitian:

1. Hasil estimasi dengan MLE menyebabkan ragam menjadi lebih besar.
2. Selang kepercayaan cenderung menjadi lebih besar karena nilai *standard error* yang besar, sehingga besar kemungkinan keputusan yang diperoleh adalah terima H_0.
3. Statistik uji t tidak menunjukkan keputusan yang signifikan, hal ini dikarenakan nilai *standard error* yang besar. Berikut persamaan untuk menperoleh nilai statistik uji t:

$$t = \frac{\hat{\beta}_j}{se\left(\hat{\beta}_j\right)}$$

   Apabila nilai *standard error* besar maka nilai statistik uji t yang dihasilkan akan bernilai kecil. Ketika nilai statistik uji t bernilai kecil dan dibandingkan dengan nilai titik kritis maka besar kemungkinan keputusan yang dihasilkan adalah terima H_0.
4. Hasil koefisien determinasi (R^2) sangat tinggi, namun nilai statistik uji t tidak memberikan keputusan yang signifikan.

2.4.2. Mengatasi Multikolinieritas

Beberapa cara yang dapat dilakukan untuk mengatasi masalah multikolinieritas menurut Draper dan Smith (1992):

1. Menambah banyak sampel
   Penambahan sampel baru memungkinkan dapat mengatasi masalah multikolinieritas.
2. Membuang satu variabel prediktor
   Ketika terjadi masalah multikolinieritas pasti terdapat satu atau beberapa variabel prediktor yang berperan besar dalam menyebabkan hubungan linier antar variabel tersebut. Salah satu langkah yang dapat dilakukan yaitu dengan membuang satu variabel prediktor yang memiliki peran besar dalam menyumbang multikolinieritas tersebut.
3. Regresi *Ridge*
   Sesuai dengan pembahasan yang dilakukan pada penelitian ini, maka salah satu cara yang dapat digunakan dalam menangani masalah multikolinieritas adalah dengan regresi *ridge*. Regresi *ridge* bekerja dengan memodifikasi matriks ***X'X*** berupa penambahan bobot (*k*).

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

2.5. Regresi *Ridge*

Regresi *ridge* dapat digunakan dalam penanganan masalah asumsi non multikolinieritas yang tidak terpenuhi, di mana terdapat korelasi yang tinggi antara variabel prediktor di dalam model sehingga menyababkan matriks $\boldsymbol{X'X}$ hampir singular dan hasil estimasi parameter menjadi kurang tepat. Metode estimasi parameter regresi *ridge* dapat menggunakan MLE dengan penambahan tetapan bias (k) sehingga koefisien menyusut mendekati nol. Menurut Draper dan Smith (1992) regresi *ridge* menghasilkan estimasi yang lebih baik daripada MLE karena menghasilkan Kuadrat Tengah Galat (KTG) yang cenderung lebih kecil. Kurtner *et al* (2004) menyatakan bahwa jika nilai (k) bernilai lebih dari nol maka koefisien regresi *ridge* bersifat bias, namun cenderung lebh stabil daripada metode estimasi dengan MLE.

Model regresi *ridge* diperoleh dengan melakukan pembakuan pada variabel respon dan variabel prediktor. Menurut Kurtner *et al* (2004) rumus pembakuan variabel adalah sebagai berikut:

$$Z_{y_i} = \frac{1}{\sqrt{n-1}}\left(\frac{y_i - \bar{y}}{S_y}\right) \quad (2.16)$$

di mana:

$\bar{y}$: rata-rata variabel respon

S_y : simpangan baku dari variabel respon

$$Z_{x_i} = \frac{1}{\sqrt{n-1}}\left(\frac{x_{ij} - \bar{x}_j}{S_x}\right) \quad (2.17)$$

di mana:

$\bar{x}_j$: rata-rata variabel prediktor ke-*j*

S_x : simpangan baku dari variabel prediktor

Setelah melakukan standarisasi variabel respon dan variabel prediktor maka diperoleh model regresi *ridge* sebagai berikut:

$$Z_{y_i} = \beta_1^* Z_{x_{i1}} + \beta_2^* Z_{x_{i2}} + \ldots + \beta_p^* Z_{x_{ip}} + \varepsilon^*$$

Sehingga diperoleh model estimasi parameter pada regresi *ridge* seperti pada persamaan (2.18):

$$\hat{Z}_{y_i} = \hat{\beta}_1^* Z_{x_{i1}} + \hat{\beta}_2^* Z_{x_{i2}} + \cdots + \hat{\beta}_p^* Z_{x_{ip}} \quad (2.18)$$

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

2.5.1. Pemilihan Tetapan Bias (*k*) pada Regresi *Ridge*

Tetapan bias (*k*) dalam regresi *ridge* menentukan besarnya hasil estimasi pada regresi *ridge*. Nilai (*k*) merupakan tetapan bias pada regresi *ridge* dengan nilai relatif kecil yang ditambahkan pada diagonal utama matriks $\boldsymbol{Z}'\boldsymbol{Z}$. Jika *k* bernilai nol maka $\hat{\boldsymbol{\beta}}_{\boldsymbol{Z}} = \hat{\boldsymbol{\beta}}$, di mana $\hat{\boldsymbol{\beta}}_{\boldsymbol{Z}}$ merupakan hasil estimasi parameter dengan MLE pada regresi *ridge* sedangkan $\hat{\boldsymbol{\beta}}$ merupakan hasil estimasi parameter dengan MLE pada regresi linier berganda (Kurtner *et al*, 2004). Ketika nilai (*k*) > 0 koefisien regresi *ridge* cenderung bias namun lebih stabil daripada metode kuadrat terkecil secara umum. Menurut Rashwan (2011), terdapat beberapa metode dalam penentuan nilai (*k*). Salah satu metode tersebut dikemukakan oleh Hoerl, Kennard dan Baldwin (1975) yang lebih dikenal dengan metode HKB, di mana dalam menentukan nilai (*k*) menggunakan persamaan sebagai berikut:

$$k = \frac{p\hat{\sigma}^2}{\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}}} \tag{2.19}$$

di mana:

p : banyak variabel prediktor

$\hat{\sigma}^2$: kuadrat tengah galat hasil dari MLE regresi linier berganda

$\hat{\boldsymbol{\beta}}$: vektor estimasi parameter dengan MLE regresi linier berganda

2.5.2. Estimasi Parameter Regresi *Ridge* dengan MLE

Metode estimasi parameter pada regresi *ridge* salah satunya dengan menggunakan MLE, karena regresi *ridge* merupakan regresi linier berganda dengan penambahan nilai (*k*) dalam metode estimasi parameter sehingga hasil estimasi parameter hampir serupa dengan regresi linier berganda dengan metode estimasi MLE maupun MKT tanpa penambahan nilai (*k*).

Menurut Kurtner *et al* (2004) pada persamaan (2.13) dapat digunakan sebagai acuan dalam menentukan fungsi *likelihood* pada regresi *ridge* sebagai berikut:

$$L\left(Z_{y_i}|Z,\beta^*,\sigma^2\right) = \prod_{i=1}^{n} f\left(Z_{y_i}|Z,\beta^*,\sigma^2\right) + k\sum_{j=1}^{p}\beta^{*2}$$

$$L\left(Z_{y_i}|Z,\beta^*,\sigma^2\right) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2}\left(\left(\mathbf{Z_y} - \mathbf{Z\beta}^*\right)'\left(\mathbf{Z_y} - \mathbf{Z\beta}^*\right)\right)\right] + k\sum_{j=1}^{p}\beta^{*2}$$

$$L\left(Z_{y_i}|Z,\beta^*,\sigma^2\right)=\left(2\pi\sigma^2\right)^{-\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2}\left(\left(\mathbf{Z_y}-\mathbf{Z\beta}^*\right)'\left(\mathbf{Z_y}-\mathbf{Z\beta}^*\right)\right)\right]+k\boldsymbol{\beta}^{*}{'}\boldsymbol{\beta}^* \quad (2.20)$$

Berikut merupakan bentuk *ln* fungsi *likelihood* dari persamaan (2.20).

$$ln\,L\left(Z_{y_i}|Z,\beta^*,\sigma^2\right)=ln\left[\left(2\pi\sigma^2\right)^{-\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2}\left(\left(\mathbf{Z_y}-\mathbf{Z\beta}^*\right)'\left(\mathbf{Z_y}-\mathbf{Z\beta}^*\right)\right)\right]\right]+k\boldsymbol{\beta}^{*}{'}\boldsymbol{\beta}^*$$

$$ln\,L\left(Z_{y_i}|Z,\beta^*,\sigma^2\right)=ln\left(2\pi\sigma^2\right)^{-\frac{n}{2}}+\ln\left(exp\left[-\frac{1}{2\sigma^2}\left(\left(\mathbf{Z_y}-\mathbf{Z\beta}^*\right)'\left(\mathbf{Z_y}-\mathbf{Z\beta}^*\right)\right)\right]\right)+k\boldsymbol{\beta}^{*}{'}\boldsymbol{\beta}^*$$

$$ln\,L\left(Z_{y_i}|Z,\beta^*,\sigma^2\right)=-\frac{n}{2}ln\left(2\pi\right)-\frac{n}{2}\ln(\sigma^2)-\frac{1}{2\sigma^2}\left(\mathbf{Z_y}-\mathbf{Z\beta}^*\right)'\left(\mathbf{Z_y}-\mathbf{Z\beta}^*\right)+k\boldsymbol{\beta}^{*}{'}\boldsymbol{\beta}^*$$

Bentuk *ln* fungsi *likelihood* digunakan untuk membentuk persamaan normal di mana turunan pertama terhadap β sama dengan nol, sehingga diperoleh hasil estimasi parameter dengan MLE untuk regresi *ridge*.

$$\begin{aligned}
\frac{\partial ln\,L\left(Z_{y_i}|Z,\beta^*,\sigma^2\right)}{\partial\boldsymbol{\beta}^*} &= 0\\
-\frac{1}{2\sigma^2}\left(-2\mathbf{Z}'\mathbf{Z_y}+2\mathbf{Z}'\mathbf{Z}\hat{\boldsymbol{\beta}}^*\right)+2k\hat{\boldsymbol{\beta}}^* &= 0\\
-\frac{1}{2\sigma^2}\left(-2\mathbf{Z}'\mathbf{Z_y}+2\mathbf{Z}'\mathbf{Z}\hat{\boldsymbol{\beta}}^*\right)+2k\hat{\boldsymbol{\beta}}^* &= 0\\
-2\mathbf{Z}'\mathbf{Z_y}+2\mathbf{Z}'\mathbf{Z}\hat{\boldsymbol{\beta}}^*+2k\hat{\boldsymbol{\beta}}^* &= 0\\
-\mathbf{Z}'\mathbf{Z_y}+\left(\mathbf{Z}'\mathbf{Z}+k\mathbf{I}\right)\hat{\boldsymbol{\beta}}^* &= 0\\
\left(\mathbf{Z}'\mathbf{Z}+k\mathbf{I}\right)\hat{\boldsymbol{\beta}}^* &= \mathbf{Z}'\mathbf{Z_y}
\end{aligned}$$

$$\hat{\beta}^*=(\boldsymbol{Z}'\boldsymbol{Z}+k\mathbf{I})^{-1}(\boldsymbol{Z}'\boldsymbol{Z_y}) \quad (2.21)$$

2.5.3. Pemeriksaan Multikolinieritas

Hasil estimasi parameter menggunakan regresi *ridge* menyebabkan nilai VIF menjadi lebih kecil, sehingga menunjukkan hasil estimasi yang sudah bebas dari multikolinieritas. Berikut pada persamaan (2.22) merupakan persamaan untuk memeriksa nilai VIF setelah dilakukan estimasi parameter menggunakan regresi *ridge* (Kurtner *et al*, 2004):

$$\text{VIF}=(\mathbf{Z'Z}+k\mathbf{I})^{-1}\mathbf{Z'Z}(\mathbf{Z'Z}+k\mathbf{I})^{-1} \quad (2.22)$$

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

2.5.4. Transformasi Balik Model Regresi *Ridge*

Setelah diperoleh hasil estimasi regresi *ridge* $\widehat{\boldsymbol{\beta}}^{*}$ maka dikembalikan pada model awal persamaan regresi untuk mendapatkan parameter regresi yang sudah bebas dari multikolinieritas (Kurtner *et al,* 2004). Berikut persamaan yang digunakan untuk mengembalikan parameter menjadi bentuk model regresi awal seperti pada persamaan (2.23) dan (2.24):

$$\hat{\beta}_j = \frac{S_y}{S_x}(\hat{\beta}_j^{\,*}) \tag{2.23}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_{i1} - \hat{\beta}_2 \bar{X}_{i2} - \ldots - \hat{\beta}_p \bar{X}_{ip} \tag{2.24}$$

2.6. Metode Bayesian Pada Regresi *Ridge*

Metode Bayesian merupakan suatu metode yang dapat menggabungkan antara distribusi prior dan informasi dari data sampel sehingga diperoleh distribusi posterior, sedangkan distribusi posterior digunakan untuk memperbarui distribusi prior dari data sampel tersebut (Pereira, 1999).

Metode Bayesian dapat digunakan sebagai salah satu metode dalam estimasi parameter. Metode Bayesian menggunakan peluang distribusi dalam mengestimasi parameter. Menurut Walpole (1995), metode Bayesian adalah suatu metode yang memanfaatkan data sampel dari suatu populasi dengan mempertimbangkan suatu distribusi awal. Jika dikaitkan dengan pendugaan parameter pada regresi *ridge* maka erat kaitannya dengan metode Bayesian pada regresi linier. Metode Bayesian pada regresi linier mengasumsikan bahwa parameter β dan σ^2 sebagai variabel yang bersifat acakdan memiliki suatu distribusi, di mana variabel *X* dan *Y* bersifat *fixed* atau sudah ditetapkan.

2.6.1. Distribusi Prior dan Posterior

Berdasarkan metode Bayesian parameter diperlakukan sebagai variabel acak yang memiliki distribusi, yaitu distribusi prior. Distribusi prior merupakan informasi awal yang digunakan dalam membentuk distribusi posterior. Distribusi prior yang dapat diterapkan pada regresi *ridge* salah satunya adalah distribusi prior yang berdasarkan pola distribusi data sampel.

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

Menurut Box dan Tiao (1973) terdapat dua jenis distribusi prior yang berdasarkan pola distribusi data sampel, yaitu:

a. *Conjugate* prior, merupakan distribusi prior dengan pola yang bergantung pada pembentukan fungsi *likelihood*.
b. *Non Conjugate* prior, merupakan prior dengan tidak bergantung pada pembentukan fungsi *likelihood*.

Metode Bayesian merupakan dasar dari statistika inferensia. Dalam teorema Bayesian menyatakan bahwa distribusi peluang untuk θ posterior data y adalah perkalian disribusi θ prior dari data sampel dan fungsi *likelihood* untuk θ yang dihasilkan dari sampel. Fungsi peluang bersyarat bagi θ jika nilai pengamatan y diketahui disebut distribusi posterior. Persamaan fungsi bersyarat bagi θ ditulis sebagai berikut:

$$f(\theta|y) = \frac{f(\theta, y)}{f(y)} \tag{2.25}$$

Persamaan (2.26) berarti bahwa fungsi peluang bersyarat satu variabel acak (θ) jika diketahui nilai variabel random yang kedua adalah fungsi peluang bersama dua variabel acak tersebut ($f(\theta, y)$) dibagi dengan fungsi peluang marginal variabel acak kedua.

$$f(\theta, y) = f(y|\theta)\, f(\theta) \tag{2.26}$$

$$f(y) = \int_{-\infty}^{\infty} f(\theta, y)\, d\theta$$

$$f(y) = \int_{-\infty}^{\infty} f(\theta)\, f(y|\theta)\, d\theta \tag{2.27}$$

Dari persamaan (2.27) dan (2.28) maka dapat ditulis persamaan posterior bagi variabel acak yang bersifat kontinyu, persamaan posterior sebagai berikut:

$$f(\theta|y) = \frac{f(y|\theta)\, f(\theta)}{\int_{-\infty}^{\infty} f(\theta)\, f(y|\theta)\, d\theta}$$

$$f(\theta|y) \propto f(y|\theta)\, f(\theta) \tag{2.28}$$

di mana:

$f(\theta|y)$: distribusi posterior
$f(y|\theta)$: fungsi *likelihood* data sampel
$f(\theta)$: distribusi prior
y : data sampel

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

θ : parameter

Regresi *ridge* berhubungan erat dengan metode Bayesian pada regresi linier, sehingga dalam penentuan distribusi prior dan posterior mengacu pada regresi linier dengan data yang berdistribusi normal. Metode Bayesian pada regresi linier mengasumsikan parameter β dan σ^2 sebagai variabel acak di mana variabel *X* dan *Y* sudah ditetapkan. Fungsi prior *conjugate* diterapkan dalam penelitian ini, di mana dalam penentuan prior berdasarkan distribusi *likelihood* yang digunakan. Nilai dari tiap parameter diperoleh sesuai dengan hasil pendugaan menggunakan metode MLE.

Menurut Greenberg (2008) *conjugate* prior menggunakan dua distribusi yang berbeda dalam keluarga yang sama ketika memiliki bentuk yang sama dan parameter yang berbeda. Sesuai dengan fungsi *likelihood* yang memanfaatkan distribusi normal, sehingga *conjugate* prior yang dapat digunakan untuk parameter yaitu distribusi normal-*inverse gamma*. Di mana prior untuk $\boldsymbol{\beta}_{\mathbf{z}}|\sigma^2$ berdistribusi normal sedangkan untuk σ^2 berdistribusi *inverse gamma* (*IG*).

$$f\left(\beta^*,\sigma^2\right)= f\left(\boldsymbol{\beta}^*|\sigma^2\right) f(\sigma^2)$$

$$f\left(\beta^*,\sigma^2\right)= N(\boldsymbol{\beta}^* \mid \boldsymbol{\beta}_{\mathbf{0}},\sigma^2\mathbf{B}_{\mathbf{0}})IG(\sigma^2 \mid \frac{\alpha_0}{2},\frac{\delta_0}{2})$$

Nilai dari $\alpha_0,\delta_0,\boldsymbol{\beta}_{\mathbf{0}},\mathbf{B}_{\mathbf{0}}$ diasumsikan diketahui. Dalam menentukan nilai dari parameter diatas dengan memanfaatkan distribusi *conjugate* prior dan beberapa komponen penyusun distribusi prosterior. Pembentukan distribusi posterior seperti pada persamaan (2.29):

$$f\left(\beta^*,\sigma^2|\mathrm{Z}_{\mathbf{y}},Z\right)\propto f(\mathbf{Z}_{\mathbf{y}} \mid \boldsymbol{\beta}^*,\sigma^2) f(\boldsymbol{\beta}^* \mid \sigma^2) f(\sigma^2)$$

$$f\left(\beta^*,\sigma^2|\mathrm{Z}_{\mathbf{y}},Z\right)\propto\left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2}\left(\mathbf{Z}_{\mathbf{y}}-\mathbf{Z}\boldsymbol{\beta}^*\right)'\left(\mathbf{Z}_{\mathbf{y}}-\mathbf{Z}\boldsymbol{\beta}^*\right)\right]$$

$$\times\left(\frac{1}{\sigma^2}\right)^{\frac{p}{2}} exp\left[-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta}^*-\boldsymbol{\beta}_0\right)'\boldsymbol{B}_0^{-1}\left(\boldsymbol{\beta}^*-\boldsymbol{\beta}_0\right)\right]$$

$$\times\left(\frac{1}{\sigma^2}\right)^{\frac{\alpha_0}{2+1}} exp\left[-\frac{\delta_0}{2\sigma^2}\right]$$

repository.ub.ac.id

$$=\left(\frac{1}{\sigma^2}\right)^{\frac{(n+\alpha_0)}{2+1}}\left(\frac{1}{\sigma^2}\right)^{\frac{p}{2}}\times exp\left[-\frac{1}{2\sigma^2}\left\{\begin{matrix}\left(\boldsymbol{Z}_y-\boldsymbol{Z}\boldsymbol{\beta}^*\right)'\left(\boldsymbol{Z}_y-\boldsymbol{Z}\boldsymbol{\beta}^*\right)\\ +\left(\boldsymbol{\beta}^*-\boldsymbol{\beta}_0\right)'\boldsymbol{B}_0^{-1}\\ \left(\boldsymbol{\beta}^*-\boldsymbol{\beta}_0\right)+\delta_0\end{matrix}\right\}\right] \tag{2.29}$$

(Greenberg, 2008)

Persamaan (2.29) dapat disederhanakan dalam bentuk sebagai berikut:

$$f(\beta^*,\sigma^2\mid Z_y)\propto\left(\frac{1}{\sigma^2}\right)^{\frac{p}{2}} exp\left[-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta}^*-\bar{\boldsymbol{\beta}}^*\right)'\mathbf{B}_1^{-1}\left(\boldsymbol{\beta}^*-\bar{\boldsymbol{\beta}}^*\right)\right]$$

$$\times\left(\frac{1}{\sigma^2}\right)^{\frac{\alpha_1}{2+1}}\exp\left[-\frac{\delta_1}{2\sigma^2}\right] \tag{2.30}$$

(Greenberg, 2008)

di mana:

$\mathbf{B}_1=\left(\mathbf{Z'Z}+\mathbf{B}_0^{-1}\right)^{-1}$

$\bar{\boldsymbol{\beta}}_z=\mathbf{B}_1(\mathbf{Z'Z_y}+\mathbf{B}_0^{-1}\boldsymbol{\beta}_0)$

$\alpha_1=\alpha_0+n$

$\delta_1=\delta_0+\mathbf{Z_y'Z_y}+\boldsymbol{\beta}_\mathbf{0}'\mathbf{B_0}^{-1}-\bar{\boldsymbol{\beta}}^*{}'\mathbf{B_1}^{-1}\bar{\boldsymbol{\beta}}^*$

Berdasarkan persamaan (2.30) maka diketahui bahwa unsur pertama persamaan tersebut merupakan fungsi kepekatan peluang untuk *p*-dimensi distribusi normal di mana $\boldsymbol{\beta}^*, N(\bar{\boldsymbol{\beta}},\sigma^2\mathbf{B_1})$ dan unsur kedua persamaan tersebut merupakan fungsi kepekatan peluang *inverse gamma* untuk σ^2, $IG\left(\frac{\alpha_1}{2},\frac{\delta_1}{2}\right)$ sehingga fungsi kepekatan peluang dari σ^2 adalah sebagai berikut:

$$f\left(\sigma^2\mid Z_y\right)=IG\left(\frac{\alpha_1}{2},\frac{\delta_1}{2}\right) \tag{2.31}$$

Distribusi posterior marginal untuk $\boldsymbol{\beta}^*$ dapat diperoleh dengan melakukan integral pada distribusi posterior $f\,(\beta^*, \sigma^2 \mid Z_y)$ terhadap σ^2.

$$f\left(\beta^*\mid Z_y\right)=\int f\left(\beta^*,\sigma^2\mid Z_y\right)d\sigma^2$$

UNIVERSITAS BRAWIJAYA

$$= \int \left(\frac{1}{\sigma^2}\right)^{\frac{(p+\alpha_1)}{2+1}} exp\left[-\frac{Q}{2\sigma^2}\right] d\sigma^2$$

di mana:

$$Q = \delta_1 + \left(\boldsymbol{\beta}^* - \bar{\boldsymbol{\beta}}^*\right)' \mathbf{B_1}^{-1} \left(\boldsymbol{\beta}^* - \bar{\boldsymbol{\beta}}^*\right)$$

Karena hanya unsur Q yang mengandung **β**, sehingga:

$$f\left(\beta^* \mid Z_y\right) \propto Q^{-\frac{(p+\alpha_1)}{2}}$$

$$\propto \left[\delta_1 + \left(\boldsymbol{\beta}^* - \bar{\boldsymbol{\beta}}^*\right)' \mathbf{B_1}^{-1} \left(\boldsymbol{\beta}^* - \bar{\boldsymbol{\beta}}^*\right)\right]^{-\frac{(p+\alpha_1)}{2}}$$

$$\propto \left[1 + \frac{1}{\alpha_1}\left(\boldsymbol{\beta}^* - \bar{\boldsymbol{\beta}}^*\right)' \left[\left(\delta_1 / \alpha_1\right) \mathbf{B_1}^{-1}\right]^{-1} \left(\boldsymbol{\beta}^* - \bar{\boldsymbol{\beta}}^*\right)\right]^{-\frac{(p+\alpha_1)}{2}} \quad (2.32)$$

(Greenberg, 2008)

2.6.2. *Markov Chain Monte Carlo* (MCMC)

MCMC suatu metode yang digunakan untuk melakukan simulasi terhadap data sampel. Metode simulasi MCMC akan menghasilkan data sampel dari suatu distribusi tertentu yang cara kerjanya menggunakan sifat dari rantai Markov yang digunakan untuk mendapatkan distribusi posterior pada suatu metode Bayesian secara akurat (Ntzoufras, 2009). Metode iterasi dengan MCMC memberikan hasil yang tergantung dengan hasil iterasi sebelumnya. MCMC merupakan suatu proses stokastik yang dapat membentuk suatu persamaan sebagai berikut:

$$f(\theta^{(t+1)} \mid \theta^{(t)}, \ldots, \theta^{(T)}) = f(\theta^{(t+1)} \mid \theta^{(t)}) \quad (2.33)$$

Distribusi posterior $f(\theta|y)$ diperoleh dengan membangkitkan dua sifat rantai markov $f(\theta^{(t+1)}|\theta^{(t)})$. Berdasarkan sifat tersebut, maka simulasi MCMC dapat diterapkan ada regresi *ridge* dengan langkah sebagai berikut (Ntzoufras, 2009):

1. Menentukan nilai awal $\left(\boldsymbol{\beta}^*, \sigma^2\right)^0$.
2. Pembangkitan data sampel sebanyak T sampai kesetimbangan distribusi tercapai.

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

3. Pemeriksaan konvergensi, jika belum diperoleh hasil yang menunjukkan konvergensi terpenuhi maka langkah pertama diulangi dengan penambahan banyak pengamatan.
4. Sebanyak ***B*** pengamatan awal dibuang.
5. Menganggap $\left\{\left(\boldsymbol{\beta}^{*},\sigma^{2}\right)^{(B+1)},\left(\boldsymbol{\beta}^{*},\sigma^{2}\right)^{(B+2)},\ldots,\left(\boldsymbol{\beta}^{*},\sigma^{2}\right)^{(T)}\right\}$ sebagai data sampel untuk distribusi posterior.
6. Memperoleh ringkasan dari distribusi posterior.

Metode MCMC memanfaatkan suatu algoritma yaitu algoritma *Gibbs Sampling*. Algoritma *Gibbs Sampling* merupakan suatu metode yang menggunakan distribusi bersyarat penuh yang digabungkan dengan distribusi posterior $f(\theta|y)$. Langkah-langkah dalam simulasi MCMC dengan menggunakan algoritma *Gibbs Sampling* adalah sebagai berikut (Ntzoufras, 2009):

1. Menentukan nilai awal $\left(\boldsymbol{\beta}^{*},\sigma^{2}\right)^{0}$.
2. Di mana t = 1, ..., T, maka langkah dalam pembangkitan sampel dengan melakukan pengulangan sebagai berikut:
   a. Menentukan nilai awal $\left(\boldsymbol{\beta}^{*},\sigma^{2}\right)=\left(\boldsymbol{\beta}^{*},\sigma^{2}\right)^{(t-1)}$.
   b. Melakukan pembaruan dengan langkah sebagai berikut:

   $\boldsymbol{\beta}^{*}{}_{1}^{(t)}$ dari $f\left(\boldsymbol{\beta}^{*}{}_{1}\,|\,\boldsymbol{\beta}^{*}{}_{2}^{(t-1)},\boldsymbol{\beta}^{*}{}_{3}^{(t-1)},\ldots,\boldsymbol{\beta}^{*}{}_{p}^{(t-1)},\left(\sigma^{2}\right)^{(t-1)},Z_{y}\right)$,

   $\boldsymbol{\beta}^{*}{}_{2}^{(t)}$ dari $f\left(\boldsymbol{\beta}^{*}{}_{2}\,|\,\boldsymbol{\beta}^{*}{}_{1}^{(t)},\boldsymbol{\beta}^{*}{}_{3}^{(t-1)},\ldots,\boldsymbol{\beta}^{*}{}_{p}^{(t-1)},\left(\sigma^{2}\right)^{(t-1)},Z_{y}\right)$,

   $\boldsymbol{\beta}^{*}{}_{3}^{(t)}$ dari $f\left(\boldsymbol{\beta}^{*}{}_{3}\,|\,\boldsymbol{\beta}^{*}{}_{1}^{(t)},\boldsymbol{\beta}^{*}{}_{2}^{(t)},\ldots,\boldsymbol{\beta}^{*}{}_{p}^{(t-1)},\left(\sigma^{2}\right)^{(t-1)},Z_{y}\right)$,

   .

   .

   .

   $\boldsymbol{\beta}^{*}{}_{p}^{(t)}$ dari $f\left(\boldsymbol{\beta}^{*}{}_{p}\,|\,\boldsymbol{\beta}^{*}{}_{1}^{(t)},\boldsymbol{\beta}^{*}{}_{2}^{(t)},\ldots,\boldsymbol{\beta}^{*}{}_{p}^{(t)},\left(\sigma^{2}\right)^{(t-1)},Z_{y}\right)$,

   $\left(\sigma^{2}\right)^{(t)}$ dari $f(\sigma^{2}{}_{p}\,|\,\boldsymbol{\beta}^{*}{}_{1}^{(t)},\boldsymbol{\beta}^{*}{}_{2}^{(t)},\ldots,\boldsymbol{\beta}^{*}{}_{p}^{(t)},Z_{y})$.

   c. Menyusun $(\boldsymbol{\beta}^{*},\sigma^{2})^{(t)}$ dan menyimpan sebagai kumpulan nilai yang dibangkitkan pada iterasi ke-(*t+1*) dari algoritma *Gibbs Sampling*.

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

Output yang dihasilkan dari algoritma *Gibbs Sampling* merupakan sampel acak yang dapat disusun pada persamaan sebagai berikut:

$$(\boldsymbol{\beta}^*,\sigma^2)^{(1)},(\boldsymbol{\beta}^*,\sigma^2)^{(2)},(\boldsymbol{\beta}^*,\sigma^2)^{(3)},\ldots,(\boldsymbol{\beta}^*,\sigma^2)^{(T)}$$

Untuk semua sampel fungsi $G(\boldsymbol{\beta z},\sigma^2)$ dengan parameter $(\boldsymbol{\beta z},\sigma^2)$, maka diperoleh hasil sebagai berikut:

1. Sampel dari parameter.

$$G((\boldsymbol{\beta}^*,\sigma^2)^{(1)}),G((\boldsymbol{\beta}^*,\sigma^2)^{(2)}),\ldots,G((\boldsymbol{\beta}^*,\sigma^2)^{(t)}),\ldots,G((\boldsymbol{\beta}^*,\sigma^2)^{(T)}) \qquad (2.34)$$

2. Ringkasan posterior berupa rata-rata dan standar deviasi untuk $G((\boldsymbol{\beta}^*,\sigma^2))$ dari sampel dengan menggunakan penduga sampel.
3. Perhitungan dan pemeriksaan korelasi antar parameter.
4. Plot dari distribusi marginal posterior.

2.7. Pemeriksaan Konvergensi

Tujuan dari pemeriksaan konvergensi yaitu untuk mengetahui pencapaian distribusi sampel dari algoritma yang digunakan pada proses simulasi. Apabila kondisi konvergen sudah tercapai maka sampel yang dihasilkan sesuai dengan distribusi target yaitu distribusi posterior. Menurut Ntzoufras (2009), terdapat beberapa metode dalam memeriksa konvergensi suatu simulasi sampel, yaitu menggunakan *trace plot*, *autocorrelation function* dan *MC Error*.

1. *Trace Plot*

Trace plot merupakan plot hasil iterasi terhadap nilai yang dibangkitkan. Konvergensi MCMC dikatakan terpenuhi apabila hasil *trace plot* tidak membentuk suatu pola tertentu. Jika tidak konvergen maka banyak iterasi perlu ditingkatkan.

2. *Autocorrelation Function* (ACF)

Apabila terdapat korelasi yag signifikan antar pengamatan secara berurutan, maka pengamatan hasil simulasi belum mencapai distribusi yang diinginkan. Plot ACF dapat dikatakan konvergen apabila *lag* pertama mendekati satu dan *lag* selanjutnya terus menurun menuju nol.

3. *MC Error*

Perhitungan *MC Error* dibagi menjadi *K batch*, apabila nilai *MC Error* kurang dari 1% simpangan baku maka kondisi konvergen MCMC terpenuhi. Langkah pertama dalam perhitungan *MC Error*

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

yaitu dengan menghitung rata-rata pengamatan hasil simulasi setiap *batch* dengan persamaan sebagai berikut:

$$\overline{G(\theta)_b} = \frac{1}{v}\sum_{t=(b-1)v+1}^{bv} G\left(\theta^{(t)}\right)$$

Setelah diperoleh nilai rata-rata pengamatan setiap *batch* maka dilakukan perhitungan rata-rata secara umum dari seluruh *batch* menggunakan persamaan berikut:

$$\overline{G(\theta)} = \frac{1}{T'}\sum_{t=1}^{T'} G\left(\theta^{(t)}\right) = \frac{1}{K}\sum_{b=1}^{K} \overline{G(\theta)_b}$$

Dari kedua rata-rata tersebut digunakan dalam perhitungan *MC Error* dengan persamaan berikut:

$$MCE[G(\theta)] = \sqrt{\frac{1}{K(K-1)}\sum_{b=1}^{K}\left(\overline{G(\theta)_b} - \overline{G(\theta)}\right)^2} \qquad (2.35)$$

di mana:

K	: banyak *batch*
v	: banyak pengamatan pada setiap *batch*
b	: indeks banyak *batch*; b=1, ..., K
t	: indeks banyak iterasi; t=1, ..., T'
$\overline{G(\theta)_b}$	: rata-rata pengamatan setiap *batch*
$\overline{G(\theta)}$	: rata-rata umum untuk seluruh *batch*

2.8. Pengujian Signifikansi Parameter

Pengujian parameter dilakukan setelah melakukan pendugaan parameter, karena dengan melakukan pengujian parameter dapat diketahui variabel prediktor yang terlibat dalam suatu model berpengaruh terhadap variabel respon atau tidak. Dalam analisis regresi linier berganda pengujian parameter dapat dilakukan secara simultan dan parsial.

2.8.1. Regresi *Ridge*

1. **Pengujian parameter secara simultan**

   Pengujian parameter secara simultan pada dasarnya menunjukkan apakah semua variabel prediktor yang dimasukkan dalam model mempunyai pengaruh secara bersama-sama terhadap variabel respon. Hipotesis yang melandasi pengujian parameter secara simultan adalah sebagai berikut:

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

$H_0: \beta_0 = \beta_1 = \beta_2 = \ldots = \beta_p = 0$ *vs*
$H_1: \exists\, \beta_j \neq 0$

Uji simultan pada parameter dapat dilihat melalui tabel ANOVA (*Analysis of Variance*) seperti pada Tabel 2.2.:

Tabel 2.2. ANOVA (*Analysis of Variance*)

SK	Db	JK	KT	Fhit
Regresi	p-1	$\text{JKR} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - \left(\frac{\mathbf{1}}{\boldsymbol{n}}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y}$	$\text{KTR} = \frac{\text{JKR}}{(\text{p}-1)}$	$\text{Fhit} = \frac{\text{KTR}}{\text{KTG}}$
Galat	n-p	$\text{JKG} = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}$	$\text{KTG} = \frac{\text{JKG}}{(\text{n}-\text{p})}$	
Total	n-1	$\text{JKT} = \mathbf{Y}'\mathbf{Y} - \left(\frac{\mathbf{1}}{\boldsymbol{n}}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y}$		

di mana:
J = matriks berisi angka 1 dengan ukuran $(n \times n)$

Kriteria hasil pengujian parameter secara simultan tolak H_0 apabila nilai dari statistik uji $F > F_{(\alpha/2;\, p-1,\, n-p)}$ .Jika H_0 ditolak berarti paling sedikit ada satu variabel prediktor yang berpengaruh secara nyata terhadap variabel respon. Sebaliknya, jika H_0 diterima berarti secara bersama-sama variabel prediktor tidak ada yang berpengaruh secara nyata terhadap variabel respon (Wathen *et al*, 2014).

2. Pengujian parameter secara parsial

Pengujian parameter secara parsial pada dasarnya menunjukkan signifikansi dari masing-masing variabel prediktor yang terlibat dalam model. Hipotesis yang melandasi pengujian parameter secara parsial adalah sebagai berikut:

$H_0: \beta_j = 0$ *vs*
$H_1: \beta_j \neq 0$

Statistik uji yang digunakan dalam pengujian parameter secara parsial adalah statistik uji *t*:

$$t = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \tag{2.36}$$

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

di mana:

j : 1, 2, ..., p

$\hat{\beta}_j$: estimasi parameter β_j

$se(\hat{\beta}_j)$: *standard error* estimasi parameter β_j

$$se(\hat{\beta}_j) = \sqrt{var(\hat{\beta}_j)} \qquad k$$

$$var(\hat{\beta}_j) = diag\left(\hat{\sigma}^2 (X'X)^{-1}\right)$$

Kriteria hasil pengujian parameter secara parsial tolak H_0 apabila nilai dari statistik uji $|t|$ > $t_{\alpha/2,\ n-p-1}$ yang berarti variabel prediktor secara individu berpengaruh secara nyata terhadap variabel respon. Sebalinya, jika H_0 diterima berarti variabel prediktor secara individu berpengaruh secara nyata terhadap variabel respon (Wathen *et al*, 2014).

2.8.2. Bayesian

Pengujian ketepatan model hasil estimasi pada metode bayesian menggunakan *Credible Interval* (CI) (Carlin dan Louis, 2009). Penggunaan CI 95% menunjukkan nilai batas bawah *percentiles* 2,5% dan nilai batas atas *percentiles* 97,5% dari nilai hasil bangkitan distribusi posterior.

Parameter dikatakan berpengaruh secara signifikan apabila selang kepercayaan tidak melewati nilai nol, sebaliknya apabila nilai estimasi parameter memiliki selang yang mengandung nol maka tidak berpengaruh secara signifikan terhadap variabel respon.

2.9. Kriteria Kebaikan Model

1. **R^2_{adj}**

Koefisien determinasi digunakan untuk mengukur nilai keragaman total variabel respon yang dapat dijelakan oleh variabel prediktor (Hestie *et al*, 2013). Sebelum melakukan perhitungan koefisien determinasi yang terkoreksi (R^2_{adj}) maka perlu diketahui nilai koefisien determinasi terlebih dahulu yang dilambangkan dengan R^2. Nilai R^2 berkisar antara nol dan satu, jika bernilai mendekati nol maka dapat dikatakan bahwa variabel prediktor kurang mampu dalam menjelaskan variabel respon.

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

Berikut merupakan persamaan untuk menentukan nilai R^2:

$$R^2 = \frac{JKR_z}{JKT_z} = 1 - \frac{JKG_z}{JKT_z}$$

di mana:
JKRz : jumlah kuadrat regresi hasil regresi *ridge*
JKTz : jumlah kuadrat total hasil regresi *ridge*
JKGz : jumlah kuadrat galat hasil regesi *ridge*

Semakin bertambah jumlah variabel prediktor maka akan menyebabkan nilai R^2 menjadi besar, sehingga diperlukan penyesuaian agar hal ini tidak terjadi. Menurut Hestie *et al* (2013) maka digunakan $R^2_{Adjusted}$, berikut persamaan yang digunakan:

$$R^2_{adj} = 1 - \frac{JKG_Z \big/ (n-p)}{JKT_z \big/ (n-1)} = 1 - (1 - R^2)\left(\frac{n-1}{n-p}\right) \tag{2.37}$$

2. *Mean Square Error* (MSE)

MSE dapat digunakan untuk melihat kebaikan model, semakin kecil nilai MSE suatu hasil estimasi parameter maka dapat dikatakan model yang terbentuk sudah cukup baik (Draper dan Smith, 1992).

$$KTG_z = \frac{JKG_z}{n-p} \tag{2.38}$$

3. *Bayesian Information Criterion* (BIC)

Kriteria penggunaan BIC sesuai dengan fungsi *ln-likelihood* dan kompleksitas model diukur dengan jumlah parameter yang digunakan oleh model. Menurut Chatterjee dan Hadi (2012) berikut persamaan untuk mencari nilai BIC pada regresi *ridge*:

$$BIC = n\ln\left(JKG_z\right) - n\ln n + \ln(n)p \tag{2.39}$$

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

4. *Akaike Information Criterion* (AIC)

AIC digunakan untuk merepresentasikan kriteria kebaikan model yang berlandaskan estimasi parameter dengan fungsi *likelihood*. Menurut Chatterjee dan Hadi (2012) berikut persamaan untuk mencari nilai AIC pada regresi *ridge*:

$$AIC = n\ln\left(JKG_z\right) - n\ln n + 2p \tag{2.40}$$

2.10. Tinjauan Non Statisika

2.10.1. Rata-rata Harapan Sekolah

Rata-rata Harapan Sekolah (RHS) digunakan sebagai salah satu indikator dalam melihat tingkat pendidikan suatu daerah. RHS merupakan lama sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak pada umur tertentu di masa mendatang. RHS dihitung untuk penduduk berusia 7 tahun ke atas dan dapat digunakan untuk mengetahui kondisi pembangunan sistem pendidikan pada berbagai jenjang yang ditunjukkan dalam bentuk lama pendidikan (dalam tahun) yang diharapkan dapat dicapai oleh setiap anak. Semakin besar RHS maka tingkat pendidikan suatu daerah tersebut juga semakin tinggi.

Beberapa faktor yang mempengaruhi RHS antara lain:

a. Rata-rata Lama Sekolah (RLS)

Kualitas pendidikan setiap daerah berbeda-beda, karena setiap penduduk dalam mengenyam pendidikan formal pastilah memiliki rentang waktu yang berbeda. Rata-rata lama sekolah dapat digunakan untuk mengetahui jumlah tahun yang digunakan oleh penduduk dalam menjalani pendidikan formal. Berikut rumusan untuk menghitung rata-rata lama sekolah:

$$\mathrm{RLS} = \frac{1}{\mathrm{P}_{15+}} \sum_{\mathrm{i}=1}^{\mathrm{P}_{15+}} \left(\mathrm{LSP}_{\mathrm{i}}\right)$$

di mana:

P_{15+} : jumlah penduduk berusia 15 tahun ke atas

LSP_i : lama sekolah Penduduk ke-*i*

Angka RLS yang tinggi menunjukkan jenjang pendidikan yang pernah/ sedang dijalani oleh seseorang, sehingga jika semakin tinggi nilai RLS mengindikasikan semakin tinggi jenjang pendidikan yang ditamatkan oleh seseorang.

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

b. Angka Partisipasi Murni (APM)

APM merupakan presentase jumlah anak pada usia sekolah tertentu yang sedang bersekolah pada jenjang pendidikan yang sesuai dengan usianya. Kegunaan APM sendiri adalah untuk mengukur proporsi anak yang bersekolah tepat waktu. Perhitungan APM sebagai berikut:

- $\text{APM SD} = \frac{\text{Jumlah murid SD usia 7}-\text{12 tahun}}{\text{Jumlah penduduk usia 7}-\text{12 tahun}} \times 100\%$
- $\text{APM SMP} = \frac{\text{Jumlah murid SMP usia 13}-\text{15 tahun}}{\text{Jumlah penduduk usia 13}-\text{15 tahun}} \times 100\%$
- $\text{APM SMA} = \frac{\text{Jumlah murid SMA usia 16}-\text{18 tahun}}{\text{Jumlah penduduk usia 16}-\text{18 tahun}} \times 100\%$

Sebagai indikator daya serap sistem pendidikan terhadap penduduk usia sekolah, maka ketika APM mendekati angka 100 dapat dikatakan bahwa seluruh anak usia sekolah dapat bersekolah tepat waktu.

c. Angka Partisipasi Sekolah (APS)

Indikator dasar yang digunakan untuk melihat akses penduduk pada fasilitas pendidikan khususnya bagi penduduk usia sekolah merupakan APS. Daya serap lembaga pendidikan terhadap penduduk usia sekolah dapat diamati dari besarnya APS. Apabila diperoleh APS yang semakin tinggi berarti bahwa semakin besar jumlah penduduk yang berkesempatan mengenyam pendidikan. Persamaan dalam menentukan APS adalah sebagai berikut:

- $\text{APS SD} = \frac{\substack{\text{Jumlah penduduk usia 7}-\text{12 tahun} \\ \text{yang masih bersekolah}}}{\text{Jumlah penduduk usia 7}-\text{12 tahun}} \times 100\%$
- $\text{APS SMP} = \frac{\substack{\text{Jumlah penduduk usia 13}-\text{15 tahun} \\ \text{yang masih bersekolah}}}{\text{Jumlah penduduk usia 13}-\text{15 tahun}} \times 100\%$
- $\text{APS SMA} = \frac{\substack{\text{Jumlah penduduk usia 16}-\text{18 tahun} \\ \text{yang masih bersekolah}}}{\text{Jumlah penduduk usia 16}-\text{18 tahun}} \times 100\%$

d. Angka Melek Huruf (AMH)

AMH menunjukkan proporsi penduduk usia 15 tahun ke atas yang mempunyai kemapuan membaca dan menulis tanpa harus mengerti apa yang di baca atau ditulis, sehingga AMH menjadi indikator penting untuk melihat sejauh mana penduduk suatu daerah terbuka terhadap pengetahuan. Rumusan untuk menentukan nilai AMH adalah sebagi berikut:

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

$$\text{AMH} = \frac{p}{q} \times 100\%$$

di mana:

p : jumlah penduduk berusia 15 tahun ke atas yang dapat membaca dan menulis

q : jumlah penduduk usia 15 tahun ke atas

Tingkat melek huruf yang tinggi menunjukkan adanya sebuah sistem pendidikan dasar yang efektif dan program keaksaraan yang memungkinkan sebagian besar penduduk untuk memperoleh kemampuan menggunakan kata-kata tertulis dalam kehidupan sehari-hari dan melanjutkan pembelajaran.

e. Penduduk dengan Tingkat Pendidikan S1

Penduduk dengan lulusan terakhir S1 mengindikasikan bahwa tingkat pendidikan yang telah diselasaikan penduduk cukup tinggi, yaitu hingga jenjang S1. Dengan melihat tamatan S1 pada suatu daerah yang semakin banyak, maka dapat menyebakan AHS meningkat.

f. Indeks Pembangunan Manusia (IPM)

Terdapat tiga dimensi dasar dalam penentuan IPM pada suatu daerah, yaitu umur panjang dan hidup sehat, pengetahuan dan standar hidup yang layak. AHS merupakan salah satu faktor yang masuk pada dimensi dasar penentuan IPM yaitu pada dimensi pengetahuan. Sehingga dengan peningkatan IPM maka diharapkan dapat menyebabkan AHS meningkat.

2.10.2. Kepadatan Penduduk

Suatu wilayah yang dikatakan padat penduduk atau tidak dapat dilihat dari tingkat kepadatan penduduk. Kepadatan penduduk pada suatu daerah menunjukkan perbandingan antara jumlah penduduk dengan luas wilayah daerah tersebut, selain itu kepadatan penduduk menyatakan jumlah rata-rata penduduk pada setiap km^2. Semakin besar angka kepadatan penduduk berarti bahea semakin padat penduduk yang mendiami suatu daerah tersebut. Rumusan dalam perhitungan kepadatan penduduk adalah sebagai berikut:

$$KP = \frac{P}{A}$$

di mana:

KP : kepadatan penduduk

P : jumlah penduduk

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

A : luas wilayah

Kepadatan penduduk menjelaskan konsentrasi penduduk di suatu daerah dan sebagai acuan dalam rangka mewujudkan pemerataan dan persebaran penduduk. Berikut beberapa faktor yang mempengaruhi kepadatan penduduk suatu wilayah:

a. Luas Wilayah

Luas wilayah menjadi suatu batasan bagi daerah tertentu, di mana ketika luas wilayah suatu daerah dalam satuan kilometer persegi semakin besar maka diharapankan pemerataan jumlah penduduk dapat stabil. Sehingga ketika luas wilayah semakin besar maka akan menyebabkan kepadatan penduduk semakin meningkat.

b. Jumlah Penduduk

Beberapa penduduk yang mendiami suatu daerah tertentu disebut dengan penduduk asli daerah. Terdapat suatu indikator yang menyatakan batasan maksimal jumlah penduduk suatu daerah, hal ini berakitan dengan luas wilayah daerah tersebut. Semakin luas wilayah daerah maka semakin mungkin jumlah penduduk akan bertambah banyak. Jika jumlah penduduk mampu merata pada suatu daerah, maka akan menyebabkan kepadatan penduduk meningkat.

c. Laju Pertumbuhan Penduduk

Sebagai indikator pertumbuhan atau peningkatan jumlah penduduk pada suatu daerah dapat dilihat dari laju pertumbuhan penduduk. Laju pertumbuhan penduduk menunjukkan angka pertambahan penduduk dalam jangka waktu tertentu, serta untuk mengetahui perubahan jumlah penduduk antara dua periode waktu. salah satu metode perhitungan dengan metode geometrik sebagai berikut:

$$r = \left(\frac{P_t}{P_0}\right)^{\frac{1}{t}} - 1$$

di mana:

r : laju pertumbuhan penduduk

P_t : jumlah penduduk tahun ke-t

P_0 : jumlah penduduk awal tahun

t : periode waktu antara tahun dasar dan tahun ke-t

Terjadi pertambahan penduduk apabila nilai laju pertumbuhan penduduk lebih dari nol, namun jika jumlah penduduk pada tahun ke-t dan awal tahun < 100 maka terjadi pengurangan penduduk pada tahun ke-t dengan tahun sebelumnya.

d. Jumlah Bayi Lahir

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

Salah satu indikator kepadatan penduduk yaitu banyak penduduk yang dapat merata pada suatu daerah. Jumlah penduduk bisa sewaktu-waktu meningkat seperti dengan pertambahan jumlah bayi yang lahir. Bayi lahir dicatat menjadi penduduk ketika dia lahir dan dinyatakan selamat dari proses kelahiran. Maka dari itu kepadatan penduduk semakin meningkat dengan pertambahan bayi yang lahir.

~ halaman ini sengaja dikosongkan ~

BAB III
METODE PENELITIAN

3.1 Sumber Data

3.1.1 Data 1

Data yang digunakan pada penelitian ini adalah data sekunder yang berasal dari Badan Pusat Statistika (BPS) mengenai Angka Harapan Sekolah Kabupaten/Kota di Jawa Timur tahun 2015. Berikut beberapa variabel yang digunakan dalam penelitian:

Y : rata-rata harapan sekolah (tahun)
X_1 : rata-rata lama sekolah (tahun)
X_2 : angka partisipasi murni (%)
X_3 : angka partisipasi sekolah (%)
X_4 : angka melek huruf (%)
X_5 : presentase penduduk lulusan S1 (%)
X_6 : indeks pembangunan manusia (%)

3.1.2 Data 2

Data yang digunakan pada penelitian ini adalah data sekunder yang berasal dari Badan Pusat Statistika (BPS) mengenai Kepadatan Penduduk menurut Kecamatan di Kabupaten Tulungagung tahun 2016. Berikut beberapa variabel yang digunakan dalam penelitian:

Y : kepadatan penduduk (1000 jiwa/km^2)
X_1 : luas wilayah (km^2)
X_2 : jumlah penduduk (ribu jiwa)
X_3 : laju pertumbuhan penduduk (%)
X_4 : jumlah bayi lahir (100 jiwa)

3.2 Metode Analisis

Metode analisis yang digunakan dalam penelitian ini dibagi menjadi dua tahapan, yaitu analisis menggunakan *Maximum Likelihood Estimation* dan Metode Bayesian pada regresi *ridge*. Proses analisis data menggunakan bantuan *software R.*

3.2.1 Estimasi Parameter Regresi *Ridge* dengan MLE

Estimasi parameter dengan MLE pada regresi *ridge* dilakukan dengan beberapa langkah sebagai berikut:

1. Melakukan pengujian asumsi untuk mendeteksi adanya pelanggaran asumsi non multikolinieritas.
2. Melakukan pembakuan terhadap variabel respon dan variabel prediktor menggunakan persamaan (2.16) dan (2.17).

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

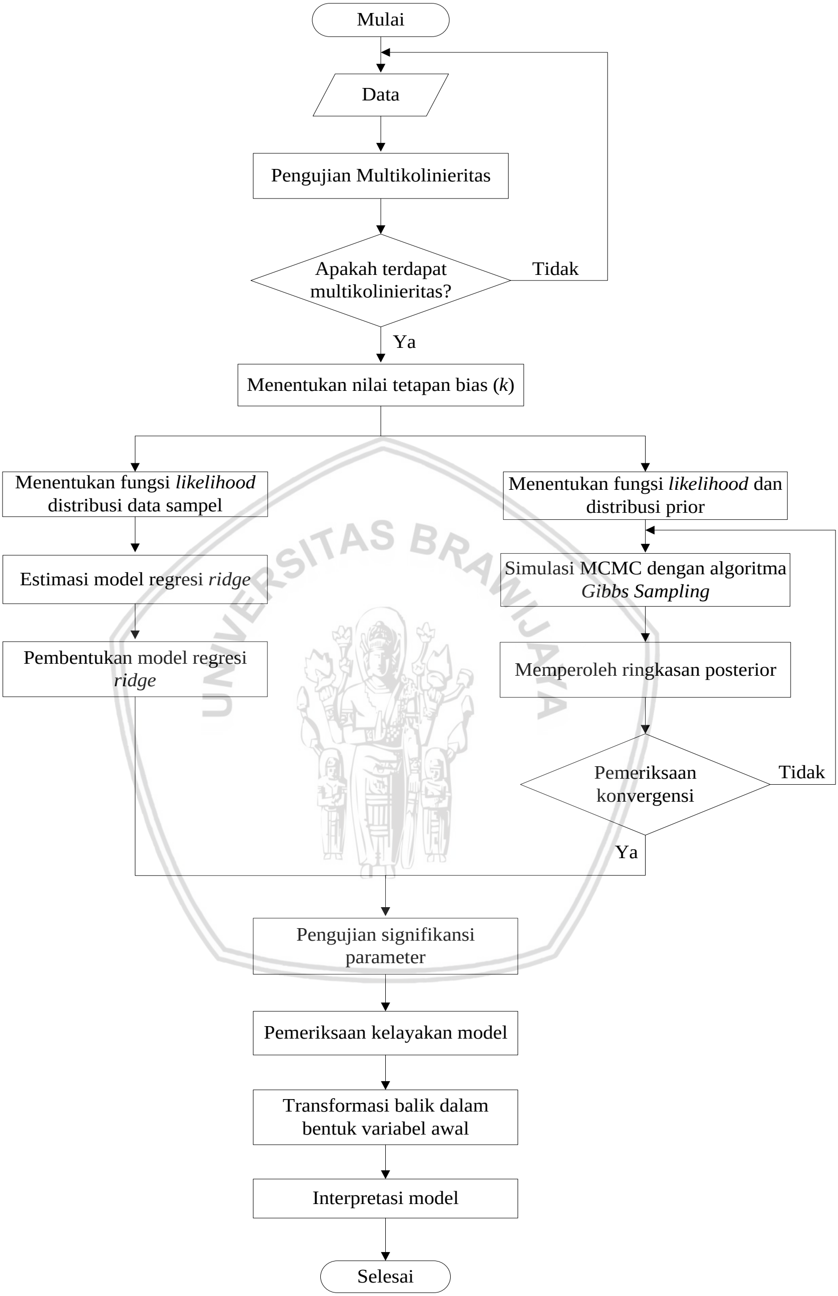
3. Menentukan nilai tetapan bias (k) sesuai dengan persamaan (2.19).
4. Menentukan estimasi paramater pada regresi *ridge* sesuai dengan persamaan (2.21).
5. Melakukan uji signifikansi parameter sesuai dengan sub bab (2.8.1).
6. Pengujian asumsi normalitas galat, homogenitas ragam galat, non autokorelasi sesuai sub bab (2.2) dan memeriksa nilai VIF akhir dengan persamaan (2.22) serta mengembalikan model regresi *ridge* pada bentuk awal model regresi berganda dengan variabel prediktor asli menggunakan persamaan (2.23) dan (2.24).

3.2.2 Estimasi Parameter Regresi *Ridge* dengan Metode Bayesian

Estimasi parameter dengan metode Bayesian pada regresi *ridge* dilakukan dengan beberapa langkah sebagai berikut:

1. Menentukan fungsi *likelihood* dan distribusi prior sesuai distribusi data sampel menggunakan persamaan (2.29).
2. Melakukan simulasi MCMC dengan algoritma *Gibbs Sampling* sesuai dengan sub bab 2.6.2.
3. Memperoleh hasil ringkasan dari distribusi posterior dari simulasi MCMC sesuai persamaan (2.34).
4. Melakukan uji signifikansi parameter sesuai dengan sub bab (2.8.2).
7. Mengembalikan model hasil estimasi *ridge* dengan metode Bayesian pada bentuk awal model regresi berganda dengan variabel prediktor asli menggunakan persamaan (2.23) dan (2.24).
8. Membandingkan metode estimasi parameter regresi *Ridge* dan Bayesian dengan kriteria kebaikan model sesuai sub bab persamaan (2.9).

Proses estimasi pada regresi *ridge* dapat dilihat pada Gambar 3.1.



Gambar 3.1 Diagram Alir Estimasi Parameter Regresi *Ridge*

~ halaman ini sengaja dikosongkan ~

BAB IV
HASIL DAN PEMBAHASAN

4.1. Data 1

Rata-rata harapan sekolah (RHS) pada suatu daerah dihitung untuk penduduk berusia 7 tahun ke atas dan dapat digunakan untuk mengetahui kondisi pembangunan sistem pendidikan pada berbagai jenjang yang dinyatakan dalam bentuk lama pendidikan (dalam tahun) yang diharapkan dapat dicapai oleh setiap anak. Semakin besar RHS maka tingkat pendidikan suatu daerah tersebut juga semakin tinggi. RHS sebagai variabel respon menunjukkan angka yang terus meningkat pada setiap Kota/ Kabupaten, peningkatan RHS dapat dilihat pada dua tahun terakhir yaitu pada tahun 2014 dan tahun 2015. Beberapa variabel yang mempengaruhi RHS yaitu rata-rata lama sekolah, angka partisipasi murni, angka partisipasi sekolah, angka melek huruf, penduduk dengan tingkat pendidikan S1, dan indeks pembangunan manusia. Terdapat salah satu variabel yang menunjukkan penurunan pada beberapa Kota/ Kabupaten yaitu variabel angka melek huruf. Beberapa Kota/ Kabupaten yang mengalami penurunan dari dua tahun terakhir dapat dilihat dari penggambaran grafik pada Lampiran 11.

4.1.1. Pemeriksaan Multikolinieritas

Pengujian asumsi non multikolinieritas pada data Rata-rata Harapan Sekolah Kota/ Kabupaten di Jawa Timur pada tahun 2015 dapat dilihat pada Tabel 4.1.:

Tabel 4.1. Nilai VIF Data 1

Variabel Prediktor	Nilai VIF	Variabel Prediktor	Nilai VIF
X_1	69.9131	X_4	14.2630
X_2	10.5189	X_5	13.2142
X_3	11.6125	X_6	26.6815

Seluruh variabel menunjukkan nilai VIF yang lebih dari 10, yang berarti asumsi tidak terpenuhi. Nilai VIF dari setiap variabel diperoleh dari hasil meregresikan masing-masing variabel prediktor dengan variabel prediktor yang lain. Asumsi non multikolinieritas yang tidak terpenuhi menunjukkan bahwa terdapat korelasi yang

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

tinggi antara keenam variabel prediktor yang mempengaruhi angka harapan sekolah tersebut.

4.1.2. Regresi *Ridge*

Salah satu metode dalam menangani masalah asumsi non multikolinieritas yang tidak terpenuhi adalah regresi *ridge* dengan memodifikasi matriks $\boldsymbol{X'X}$, di mana matriks tersebut ditambahkan nilai tetapan bias (k). Estimasi menggunakan regresi *ridge* berdasarkan data yang telah dibakukan, kemudian hasil estimasi dengan data yang dibakukan selanjutnya dikembalikan pada bentuk regresi awal yang sudah terbebas dari multikolinieritas.

1. Estimasi Parameter Regresi *Ridge*

Dari hasil perhitungan, maka diperoleh nilai tetapan bias (k) sebesar 1.1784. Berikut merupakan model regresi dengan variabel respon dan prediktor yang dibakukan menggunakan MLE:

$$\hat{Z}_y = 0.1119\, Z_1 + 0.0889\, Z_2 + 0.1345\, Z_3 + 0.0213\, Z_4 + 0.2169\, Z_5 + 0.1484\, Z_6 \quad (4.1)$$

2. Uji Signifikansi Parameter Regresi *Ridge*

a. Pengujian Parameter Secara Simultan

Pengujian signifikansi secara simultan digunakan untuk mengetahui variabel prediktor yang terlibat dalam penelitian berpengaruh secara bersama-sama terhadap variabel respon atau tidak.

H_0: $\beta_1 = \beta_2 = \cdots = \beta_p = 0$ *vs*

H_1: $\exists\, \beta_j \neq 0$

Tabel 4.2. menunjukkan hasil uji simultan parameter regresi *ridge* yang disajikan dalam bentuk tabel ANOVA.

Tabel 4.2. Hasil Uji Simultan Data 1

SK	Db	JK	KT	Fhit
Regresi	5	0.5664	0.1132	8.3613
Galat	32	0.4336	0.0135	
Total	37	1		

Dalam pengambilan keputusan uji simultan diketahui nilai Fhit sebesar 8.3613 ≥ $\text{Ftabel}_{(0.025;5;32)}$ (2.995) dengan taraf nyata (α = 0.05), sehingga H_0 ditolak yang berarti bahwa paling sedikit terdapat satu variabel prediktor yang berpengaruh secara nyata terhadap Rata-rata Harapan Sekolah pada Kota/ Kabupaten di Jawa Timur.

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

b. Pengujian Parameter Secara Parsial

Selain dilakukan pengujian signifikansi secara simultan dilakukan juga uji parsial untuk mengetahui variabel prediktor yang berpengaruh nyata secara parsial atau individu.

H_0: $\beta_j = 0$ *vs*

H_1: $\beta_j \neq 0$

Hasil pengujian parameter disajikan pada Tabel 4.3., dengan nilai $t_{tabel\ (0.025;32)} = 2.037$.

Tabel 4.3. Hasil Uji Parsial Data 1

j	**Nilai Estimasi**	**se($\hat{\beta}_j$)**	**Statistik Uji t**	***p-value***	**Keputusan**
1	0.1119	0.0247	4.5341	6.2×10^{-5}	Tolak H_0
2	0.0889	0.0390	2.2761	0.0288	Tolak H_0
3	0.1345	0.0356	3.7469	0.0006	Tolak H_0
4	0.0213	0.0397	0.5360	0.5952	Terima H_0
5	0.2169	0.0405	5.3559	5.04×10^{-6}	Tolak H_0
6	0.1484	0.0265	5.5957	2.41×10^{-6}	Tolak H_0

Dari uji parsial maka dapat diketahui bahwa rata-rata lama sekolah, angka partisipasi murni, angka partisipasi sekolah, banyak lulusan S1, dan indeks pembangunan manusia secara idividu memberikan pengaruh secara nyata terhadap Rata-rata Harapan Sekolah di Kota/Kabupaten Jawa Timur tahun 2015.

3. Kelayakan Model

a. Pemeriksaan Multikolinieritas Akhir

Hasil penanganan asumsi non multikolinieritas yang tidak terpenuhi dengan menggunakan regesi *ridge* maka diperoleh hasil estimasi yang sudah terbebas dari masalah multikolinieritas.

Tabel 4.4. Nilai VIF Akhir Data 1

Variabel Prediktor	**Nilai VIF**	**Variabel Prediktor**	**Nilai VIF**	**Variabel Prediktor**	**Nilai VIF**
Z_1	0.0449	Z_3	0.0908	Z_5	0.1211
Z_2	0.1125	Z_4	0.1160	Z_6	0.0519

Berdasarkan Tabel 4.4. pemeriksaan multikolinieritas dilakukan dengan menghitung nilai VIF setelah dilakukan regresi

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

ridge yang memanfaatkan data hasil pembakuan dengan penambahan nilai (k) seperti pada persamaan (2.22). Nilai VIF yang kurang dari sepuluh menunjukkan bahwa masalah multikolinieritas sudah teratasi.

b. Non Autokorelasi

Pengujian asumsi non autokorelasi dilakukan dengan menggunakan statistik uji *Durbin Watson* (DW) dengan hipotesis sebagai berikut:

H_0: $\rho = 0$ *vs*

H_1: $\rho \neq 0$

Hasil statistik DW dengan *p-value* sebesar 0.7118 yang menunjukkan lebih besar dari taraf nyata (α) = 0.05 sehingga dapat disimpulkan bahwa terima H_0 yang berarti asumsi non autokorelasi terpenuhi. Selain dengan melihat *p-value* maka dapat dilakukan dengan cara lain yaitu dengan melihat nilai statistik uji DW sebesar 2.1779 diperoleh nilai *dU* (1.1463) < *d* (2.2784) < *4-dU* (2.8537) sesuai dengan tabel DW, sehingga dapat diambil kesimpulan bahwa terima H_0 yang berarti asumsi non autokorelasi terpenuhi.

c. Homogenitas Ragam Galat

Pengujian asumsi homogenitas ragam galat dengan menggunakan statistik uji *Breusch Pagan* dengan hipotesis sebagai berikut:

H_0: $\alpha_0 = \alpha_1 = \ldots = \alpha_p$ *vs*

H_1: $\alpha_0 \neq \alpha_1 \neq \ldots \neq \alpha_p$

Hasil uji *Breusch Pagan* dengan *p-value* sebesar 0.2240 yang menunjukkan lebih besar dari taraf nyata (α) = 0.05, sehingga dapat disimpulkan bahwa terima H_0 yang berarti asumsi homogenitas ragam galat terpenuhi. Selain dengan melihat *p-value* maka dapat dilakukan dengan cara lain yaitu dengan melihat nilai statistik uji LM sebesar (8.1975) < $\chi^2_{(0.025;6)}$(14.45) yang berarti keputusan terima H_0, sehingga dapat disimpulkan bahwa asumsi homogenitas ragam galat terpenuhi.

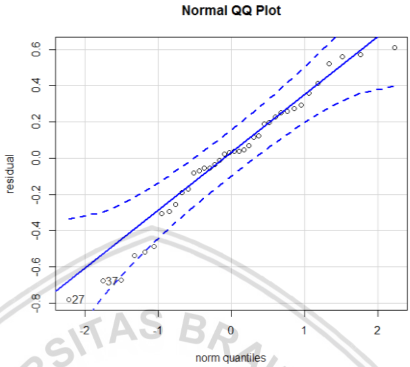
repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

d. Normalitas Galat

Asumsi normalitas galat menunjukkan pola sebaran dari galat data sampel, hipotesis yang digunakan adalah sebagai berikut:

H_0: $\varepsilon \sim N\ (\mu, \sigma^2)$ *vs*

H_1: $\varepsilon \nsim N\ (\mu, \sigma^2)$



Gambar 4.1. Uji Normalitas Galat Data 1

Gambar 4.1. merupakan pola sebaran nilai galat yang berada disekitar garis regresi, hal ini menunjukkan bahwa galat menyebar secara normal. Selain dengan melihat hasil plot nilai galat, maka dapat dilakukan dengan melihat *p-value* statistik uji *Kolmogorov Smirnov* sebesar 0.1696 yang menunjukkan lebih besar dari taraf nyata (α) = 0.05, sehingga dapat disimpulkan bahwa terima H_0 yang berarti galat berdistribusi normal.

e. Transformasi Balik Model Regresi *Ridge*

Transformasi balik pada model regresi *ridge* dilakukan jika model tersebut sudah terbebas dari multikolinieritas sesuai dengan persamaan (2.23) dan (2.24), maka diperoleh model regresi yang telah ditransformasi balik sebagai berikut:

$$\hat{Y} = 6.5642 + 0.0567X_1 + 0.0134X_2 + 0.0265X_3 + 0.0035X_4 + 0.0689X_5 + 0.0239X_6 \quad (4.2)$$

4.1.3. Regresi *Ridge* Bayesian

Sebagai salah satu metode estimasi parameter berupa estimasi interval, metode bayesian dapat diterapkan pada proses estimasi parameter regresi *ridge*, karena pada metode Bayesian tidak mengharuskan suatu asumsi terpenuhi. Metode bayesian dalam estimasi parameter memanfaatkan distribusi awal dari data sampel

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

yang disebut distribusi prior, kemudian dari distribusi prior ini digabungkan dengan fungsi *likelihood* sesuai data sampel sehingga terbentuk distribusi posterior yang mendasari metode estimasi pada metode Bayesian.

1. Pembentukan Fungsi *Likelihood* dan Distribusi Prior

Pembentukan fungsi *likelihood* memerlukan informasi mengenai distribusi data sampel yang digunakan, sesuai dengan pengujian asumsi menunjukkan bahwa galat dari data memiliki distribusi normal. Fungsi *likelihood* untuk regresi *ridge* dengan distribusi normal sesuai dengan persamaan (2.20) adalah sebagai berikut:

$$L(\boldsymbol{\beta}_{\boldsymbol{z}},\sigma^2|\boldsymbol{Y}^*,\boldsymbol{Z}) = (2\pi\sigma^2)^{-\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2}\left((\boldsymbol{Y}^*-\boldsymbol{Z}\boldsymbol{\beta}_{\boldsymbol{z}})'(\boldsymbol{Y}^*-\boldsymbol{Z}\boldsymbol{\beta}_{\boldsymbol{z}})+k\boldsymbol{\beta}'\boldsymbol{\beta}\right)\right]$$

Setelah diperoleh fungsi *likelihood*, selanjutnya yaitu membentuk distribusi prior. Di mana pada penelitian ini menggunakan *conjugate* prior dengan pendekatan *pseudo* prior. *Conjugate* prior menggunakan distribusi eksponensial dalam satu keluarga, sedangkan *pseudo* prior memanfaatkan hasil estimasi sebelumnya. Hasil estimasi sebelumnya menggunakan MLE sehingga dapat dimanfaatkan dalam penentuan distribusi prior pada metode estimasi Bayesian ini. Fungsi *likelihood* dan distribusi prior yang sudah terbentuk maka selanjutnya digunakan untuk menentukan distribusi posterior.

2. Estimasi Parameter *Ridge* Bayesian

Metode Bayesian sebagai salah satu metode estimasi yang memanfaatkan distribusi dari data sampel menghendaki terpenuhinya suatu kondisi dimana hasil estimasi menunjukkan hasil yang konvergen. Konvergensi menunjukkan hasil estimasi tersebut sudah mendekati hasil yang sebenarnya, pada penelitian ini yaitu mendekati hasil estimasi dengan menggunakan regresi *ridge*. Proses estimasi dilakukan dengan simulasi sampel sebanyak 90000 kali yang memanfatkan metode metode *Gibbs Sampling*.

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

Tabel 4.5. Hasil Estimasi Metode Bayesian Data 1

j	Nilai Estimasi	*Credible Interval*	
		2.5%	97.5%
1	0.1176	0.0700	0.1660
2	0.0995	0.0270	0.1700
3	0.1515	0.0850	0.2180
4	0.0065	-0.1140	0.1260
5	0.2579	0.1830	0.3340
6	0.1592	0.1080	0.2100

Berdasarkan Tabel 4.5. menunjukkan bahwa terdapat satu variabel yang tidak berpengaruh, karena nilai *Credible Interval* (CI) menunjukkan nilai yang melewati nol. Pengujian signifikansi pada hasil estimasi dengan metode Bayesian menunjukkan hasil yang sama dengan MLE, hal ini menunjukkan bahwa metode Bayesian sesuai jika digunakan untuk mengestimasi model regresi *ridge*.

3. Pemeriksaan Konvergensi

Menurut Ntzoufras (2009) pemeriksaan konvergensi pada hasil estimasi dengan metode Bayesian dapat dilakukan dengan tiga cara yaitu dengan memeriksa nilai *MC Error*, ACF dan *Trace Plot*.

Tabel 4.6. Nilai *MC Error* Data 1

j	sd	1%*sd	*MC Error*
1	0.0243	2.4299×10^{-4}	1.2453×10^{-4}
2	0.0366	3.6684×10^{-4}	1.9414×10^{-4}
3	0.0340	3.0538×10^{-4}	1.7295×10^{-4}
4	0.0602	6.0281×10^{-4}	3.4485×10^{-4}
5	0.0388	3.8832×10^{-4}	2.1967×10^{-4}
6	0.0261	2.6147×10^{-4}	1.4849×10^{-4}

Nilai *MC Error* dikatakan konvergen apabila kurang dari 1% nilai *standard deviasi*, pada Tabel 4.6. menunjukkan bahwa nilai *MC Error* sudah sesuai sehingga dapat dikatakan hasil estimasi sudah konvergen. *Trace* plot menunjukkan sudah tidak membentuk suatu pola tertetu sesuai dengan Lampiran 5 sehingga dapat dikatakan konvergen. Plot ACF dapat dilihat pada Lampiran 6 menunjukkan bahwa *lag* pertama mendekati nilai satu dan *lag* selanjutnya bernilai mendekati nol berarti sudah dapat dikatakan konvergen.

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

Ketiga cara pemeriksaan konvergensi menunjukkan bahwa hasil estimasi dengan metode Bayesian sudah konvergen, sehingga tidak perlu dilakukan penambahan sampel dalam iterasi.

4. Transformasi Balik Model Regresi *Ridge* Bayesian

Model transformasi balik yang diperoleh adalah sebagai berikut:

$$\hat{Y} = 6.1670 + 0.0670X_1 + 0.0150X_2 + 0.030X_3 + 0.0010X_4 + 0.0820X_5 + 0.0260X_6 \quad (4.3)$$

Model di atas menunjukkan hasil nilai estimasi yang tidak jauh berbeda dengan hasil estimasi menggunakan regresi *ridge* pada persamaan (4.2), sehingga dapat disimpulkan bahwa metode Bayesian sesuai untuk mengestimasi model regresi *ridge*.

4.1.4. Interpretasi Model dan Pembahasan

Sesuai persamaan (4.2) dan (4.3) maka estimasi paraeter dengan kedua metode dapat dilihat pada Tabel 4.7, di mana hasil estimasi menunjukkan nilai yang tidak jauh berbeda.

Tabel 4.7. Hasil Estimasi Parameter Data 1

j	**Regresi *Ridge***	**Bayesian**
Intersep	6.5642	6.1670
1	0.0567	0.0670
2	0.0134	0.0150
3	0.0265	0.0300
4	0.0035	0.0010
5	0.0689	0.0820
6	0.0239	0.0260

1. Regresi *Ridge*

Hasil estimasi regresi *ridge* diketahui bahwa semua variabel prediktor memberikan pengaruh nyata terhadap angka harapan sekolah. Persamaan (4.2) menunjukkan bahwa semua variabel prediktor memberikan pengaruh positif terhadap variabel respon. Di mana setiap peningkatan 100 tahun rata-rata lama sekolah maka akan meningkatkan rata-rata harapan sekolah sebesar 5.67 tahun rata-rata harapan sekolah, setiap peningkatan 100% angka partisipasi murni maka akan meningkatkan rata-rata harapan sekolah sebesar 1.34 tahun, dan setiap peningkatan 100% angka

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

partisipasi sekolah maka akan meningkatkan rata-rata harapan sekolah sebesar 2.65 tahun dengan variabel prediktor lain dinggap konstan. Apabila terjadi peningkatan sebesar 100% angka melek huruf maka akan meningkatkan rata-rata harapan sekolah sebesar 0.35 tahun, atau untuk setiap kenaikan 10 kali lipat angka melek huruf maka akan meningkatkan rata-rata harapan sekolah sebesar 3.5 tahun, setiap peningkatan 100% banyak lulusan S1 maka akan meningkatkan rata-rata harapan sekolah sebesar 6.89 tahun, dan setiap peningkatan 100% indeks pembangunan manusia maka akan meningkatkan rata-rata harapan sekolah sebesar 2.39 tahun dengan variabel prediktor lain dinggap konstan.

Dari hasil analisis diketahui bahwa salah satu variabel prediktor tidak berpengaruh secara nyata terhadap rata-rata harapan sekolah, yaitu variabel angka melek huruf. Pada kehidupan nyata sebagian besar penduduk Jawa Timur jika merasa sudah bisa membaca dan menulis kata, maka tidak ada keinginan untuk melanjutkan sekolah ke jenjang yang lebih tinggi sehingga tidak berpengaruh terhadap tingkat angka harapan sekolah. Kondisi seperti ini dapat dilihat pada Lampiran 11 bahwa pada tahun 2014 ke tahun 2015 terdapat beberapa Kota/ Kabupaten di Jawa Timur yang tidak mengalami kenaikan pada tingkat angka melek huruf, bahkan cenderung menurun. Beberapa Kota/ Kabupaten tersebut antara lain Ponorogo, Trenggalek, Jember, Banyuwangi, Bondowoso, Pasuruan, Magetan, Ngawi, Lamongan, Pamekasan, Sumenep, Kota Probolinggo, dan Kota Pasuruan.

2. Regresi *Ridge* Bayesian

Setiap peningkatan 100 tahun rata-rata lama sekolah maka akan meningkatkan rata-rata harapan sekolah sebesar 6.7 tahun. Setiap peningkatan 100% angka partisipasi murni maka akan meningkatkan rata-rata harapan sekolah sebesar 1.5 tahun. Setiap peningkatan 100% angka partisipasi sekolah maka akan meningkatkan rata-rata harapan sekolah sebesar 3 tahun. Setiap peningkatan 100 tahun angka melek huruf maka akan meningkatkan rata-rata harapan sekolah sebesar 0.1 tahun. Setiap peningkatan 100% banyak lulusan S1 maka akan meningkatkan arata-rata harapan sekolah sebesar 8.2 tahun. Selanjutnya setiap peningkatan 100% indeks pembangunan manusia maka akan meningkatkan rata-rata harapan sekolah sebesar 2.6 tahun. Pada

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

penelitian ini variabel angka melek huruf tidak berpengaruh terhadap angka harapan sekolah.

Seluruh variabel memberikan pengaruh positif terhadap angka harapan sekolah, tetapi untuk variabel angka melek huruf tidak berpengaruh secara nyata terhadap angka harapan sekolah. Penyebab hal tersebuat yaitu ketika seseorang sudah mampu mengeja kata, maka tidak tumbuh keinginan untuk meneruskan sekolah ke jenjang yang lebih tinggi. Terbukti dari tahun 2014 sampai 2015 terdapat beberapa Kota/ Kabupaten yang mengalami penurunan pada angka melek huruf. Beberapa Kota/ Kabupaten yang mengalami penurunan angka melek huruf yaitu Kabupaten Ponorogo, Trenggalek, Tulungagung, Jember, Banyuwangi, Bondowoso, Situbondo, Pasuruan, Magetan, Ngawi, Lamongan, Gresik, Pamekasan, Sumenep, Kota Probolinggo dan Kota Pasuruan.

4.2. Data 2

Kabupaten Tulungagung merupakan salah satu daerah yang berada di propinsi Jawa Timur dengan 19 Kecamatan. Luas wilayah Kabupaten Tulungagung sebesar 1055.65 km^2, di mana dengan luas wilayah tersebuat maka dapat diketahui tingkat kepadatan penduduk pada Kabupaten tersebut. Kepadatan penduduk Kabupaten Tulungagung menjadi variabel respon pada penelitian ini, dengan beberapa variabel prediktor yaitu luas wilayah, jumlah penduduk, laju pertumbuhan penduduk dan jumlah bayi lahir. Terdapat penurunan kepadatan penduduk pada dua tahun terakhir yaitu pada tahun 2015 dan tahun 2016 pada beberapa Kecamatan di Kabupaten Tulungagung. Hal ini disebabkan oleh luas wilayah yang memiliki luasan tetap karena kecil kemungkinan suatu wilayah akan meningkat luasannya, namun disisi lain tidak terdapat sumber daya yang memadai seperti lapangan pekerjaan yang kurang, sehingga menyebabkan beberapa penduduk berpindah ke daerah lain. Selain itu variabel laju pertumbuhan penduduk mengalami kenaikan dari tahun 2015 ke tahun 2016, hal ini berbanding terbalik dengan usaha pemerintah Kabupaten Tulungagung untuk memperkecil laju pertumbuhan penduduk tiap tahunnya.

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

4.2.1. Pemeriksaan Multikolinieritas

Salah satu asumsi regresi linier berganda yaitu non multikolinieritas. Pengujian asumsi non multikolinieritas pada data Kepadatan Penduduk menurut Kecamatan di Kabupaten Tulungagung pada tahun 2016 sebagai berikut:

Tabel 4.8. Nilai VIF Data 2

Variabel Prediktor	Nilai VIF	Variabel Prediktor	Nilai VIF
X_1	1.4729	X_3	1.4952
X_2	56.3735	X_4	52.3785

Nilai VIF dari Tabel 4.8. diperoleh dari meregresikan masing-masingvariabel prediktor dengan variabel prediktor yang lain. Terdapat dua variabel prediktor yang menyebabkan asumsi non multikolinieritas tidak terpenuhi dengan nilai VIF yang lebih dari 10. Asumsi yang tidak terpenuhi ini menunjukkan bahwa terdapat korelasi yang tinggi antara banyak penduduk dan banyak bayi lahir.

4.2.2. Regresi Ridge

Regresi *ridge* digunakan sebagai salah satu metode dalam menangani masalah asumsi non multikolinieritas yang tidak terpenuhi dengan memodifikasi matriks ***X'X*** dengan menambahkan nilai tetapan bias (k). Estimasi menggunakan regresi ridge berdasarkan data yang telah distandarisasi, di mana hasil estimasi dengan data standarisasi ini selanjutnya dikembalikan pada bentuk regresi awal.

1. **Estimasi Parameter Regresi *Ridge***

   Dari hasil perhitungan nilai (k) maka diperoleh sebesar 0.0892. Berikut merupakan model regresi yang terbentuk dari hasil estimasi parameter regresi *ridge* dengan menggunakan MLE:

   $$\hat{Z}_y = -0.5768\, Z_1 + 0.2485\, Z_2 + 0.0804\, Z_3 + 0.1947\, Z_4 \quad (4.4)$$

2. **Pengujian Signifikansi Parameter Regresi *Ridge***

   a. **Pengujian Signifikansi Parameter Secara Simultan**

   Hasil estimasi parameter dengan regresi *ridge* sebagai salah satu metode yang dapat menangani asumsi non multikolinieritas perlu dilakukan pengujian signifikansi hasil estimasi secara serentak atau simultan. Pengujian signifikansi secara simultan digunakan untuk mengetahui variabel prediktor

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

yang terlibat dalam penelitian berpengaruh bersama-sama terhadap variabel respon atau tidak.

H_0: $\beta_1 = \beta_2 = \ldots = \beta_p = 0$ *vs*

H_1: $\exists\, \beta_j \neq 0$

Tabel 4.9. merupakan hasil uji simultan parameter regresi *ridge* yang disajikan dalam bentuk tabel ANOVA sebagai berikut:

Tabel 4.9. Hasil Uji Simultan Data 2

SK	db	JK	KT	Fhit
Regresi	3	0.9069	0.3023	48.7506
Galat	15	0.0930	0.0062	
Total	18	1		

Pengambilan keputusan untuk uji simultan menunjukkan nilai Fhit sebesar 48.7506 ≥ $Ftabel_{(0.025;3;15)}$ (4.15) dengan taraf nyata (α = 0.05), sehingga H_0 ditolak yang berarti bahwa paling sedikit terdapat satu variabel prediktor yang berpengaruh secara nyata terhadap Kepadatan Penduduk menurut Kecamatan di Kabupaten Tulungagung tahun 2016.

b. Pengujian Signifikansi Parameter Secara Parsial

Selain dilakukan pengujian signifikansi secara simultan dilakukan juga uji parsial untuk mengetahui variabel prediktor yang berpengaruh nyata secara parsial atau individu.

H_0: $\beta_j = 0$ *vs*

H_1: $\beta_j \neq 0$

Tabel 4.10. merupakan uji parsial parameter regresi *ridge*, dengan nilai $t_{tabel\ (0.025;32)}$ = 2.037.

Tabel 4.10. Hasil Uji Parsial Data 2

j	$\widehat{\beta}_j$	se($\widehat{\beta}_j$)	Statistik Uji t	*p-value*	Keputusan
1	-0.5768	0.0828	-6.9671	2.27×10^{-6}	Tolak H_0
2	0.2485	0.0705	3.5272	0.0026	Tolak H_0
3	0.0804	0.0770	1.0439	0.3111	Terima H_0
4	0.1947	0.0714	2.7244	0.0144	Tolak H_0

Berdasarkan uji parsial diketahui bahwa luas wilayah, banyak penduduk dan banyak bayi yang lahir secara individu

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

memberikan pengaruh secara nyata terhadap Kepadatan Penduduk menurut Kecamatan di Kabupaten Tulungagung tahun 2016.

3. Kelayakan Model

a. Pemeriksaan Multikolinieritas Akhir

Sesuai dengan hasil estimasi dengan regresi *ridge* maka diperoleh hasil estimasi yang sudah terbebas dari masalah multikolinieritas. Hal ini ditunjukkan dengan penghitungan nilai VIF sesuai persamaan (2.22) dengan variabel yang telah dibakukan dan ditambah dengan nilai (k).

Tabel 4.11. Nilai VIF Akhir Data 2

Variabel Prediktor	Nilai VIF
Z_1	1.1053
Z_2	0.8006
Z_3	0.9565
Z_4	0.8242

VIF bernilai kurang dari sepuluh yang menunjukkan tidak terdapat korelasi antar variabel prediktor yang telah dibakukan, sehingga dapat disimpulkan bahwa rergesi *ridge* mampu menangani masalah asumsi non multikolinieritas yang tidak terpenuhi.

b. Non Autokorelasi

Pengujian asumsi non autokorelasi dapat dengan mengunakan statistik uji *Durbin Watson* (DW) dengan hipotesis sebagai berikut:

H_0: $\rho = 0$ *vs*

H_1: $\rho \neq 0$

Hasil statistik DW dengan *p-value* sebesar 0.5737 yang menunjukkan lebih besar dari taraf nyata (α) = 0.05 sehingga dapat disimpulkan bahwa terima H_0 yang berarti asumsi non autokorelasi terpenuhi. Selain dengan melihat nilai *p-value* maka dapat dilakukan dengan cara lain yaitu dengan melihat nilai statistik uji DW sebesar 2.1875 diperoleh nilai 4-dU (2.1518) < d (2.1875) < 4-dL (3.1412) sesuai dengan tabel DW. Kesimpulan yang diperoleh sesuai dengan hasil tersebut adalah masih ragu-ragu,

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

sehingga dilakukan *Run Test* untuk memeriksa keacakan nilai galat. Hasil *Run Test* menunjukkan *p-value* > α (0.05) yang berarti asumsi non autokorelasi terpenuhi.

c. Homogenitas Ragam Galat

Pengujian asumsi homogenitas ragam galat dapat dengan menggunakan statistik uji *Breusch Pagan* dengan hipotesis sebagai berikut:

H_0: $\alpha_0 = \alpha_1 = \ldots = \alpha_p$ *vs*

H_1: $\alpha_0 \neq \alpha_1 \neq \ldots \neq \alpha_p$

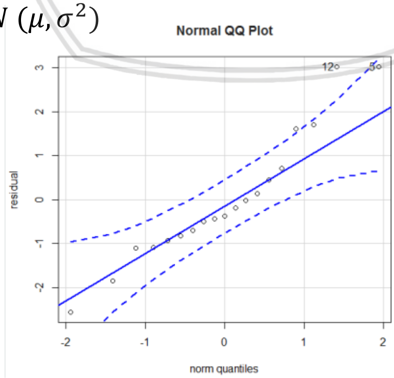
Hasil uji *Breusch Pagan* dengan *p-value* sebesar 0.9984 yang menunjukkan lebih besar dari taraf nyata (α) = 0.05, sehingga dapat disimpulkan bahwa terima H_0 yang berarti asumsi homogenitas ragam galat terpenuhi. Selain dengan melihat *p-value* maka dapat dilakukan dengan cara lain yaitu dengan melihat nilai statistik uji LM sebesar (0.1362) < $\chi^2_{(0.025;4)}$ (11.14) yang berarti keputusan terima H_0, sehingga dapat disimpulkan bahwa asumsi homogenitas ragam galat terpenuhi.

d. Normalitas Galat

Asumsi normalitas galat menunjukkan pola sebaran dari galat yang mencerminkan pola sebaran data sampel yang akan dianalisis, hipotesis yang digunakan adalah sebagai berikut:

H_0: $\varepsilon \sim N\ (\mu, \sigma^2)$ *vs*

H_1: $\varepsilon \nsim N\ (\mu, \sigma^2)$



Gambar 4.2 Uji Normalitas Galat Data 2

Gambar 4.2 merupakan pola sebaran nilai galat yang berada disekitar garis regresi, hal ini menunjukkan bahwa galat menyebar secara normal. Selain dengan melihat hasil plot nilai galat, maka dapat dilakukan dengan melihat *p-value* dari statistik uji *Kolmogorov Smirnov* sebesar 0.2813 yang menunjukkan lebih besar dari taraf nyata (α) = 0.05, sehingga dapat disimpulkan bahwa terima H_0 yang berarti galat berdistribusi normal.

e. Transformasi Balik Model Regresi *Ridge*

Transformasi balik pada model regresi *ridge* dilakukan jika model tersebut sudah terbebas dari multikolinieritas sesuai dengan persamaan (2.23) dan (2.24), maka diperoleh model regresi sebagai berikut:

$$\hat{Y} = 1.0258 - 0.0130X_1 + 0.0086X_2 + 0.1843X_3 + 0.0446X_4 \quad (4.5)$$

4.2.3. Regresi *Ridge* Bayesian

Sebagai salah satu metode estimasi parameter berupa estimasi interval, metode bayesian dapat diterapkan pada proses estimasi parameter regresi *ridge*, karena pada metode Bayesian tidak mengharuskan suatu asumsi terpenuhi. Metode bayesian dalam estimasi parameter memanfaatkan distribusi awal dari data sampel yang disebut distribusi prior, kemudian dari distribusi prior ini digabungkan dengan fungsi *likelihood* sesuai dengan data sampel sehingga terbentuk distribusi posterior yang mendasari metode estimasi pada metode Bayesian.

1. Pembentukan Fungsi *Likelihood* dan Distribusi Prior

Estimasi parameter dengan memanfaatkan metode Bayesian pada regresi *ridge* memerlukan fungsi *likelihood* dan informasi mengenai distribusi data sampel yang digunakan, sesuai dengan pengujian asumsi menunjukkan bahwa galat dari data memiliki distribusi normal. Fungsi *likelihood* untuk regresi *ridge* dengan distribusi normal sesuai dengan persamaan (2.20) adalah sebagai berikut:

$$L(\boldsymbol{\beta_z}, \sigma^2 | \boldsymbol{Y}^*, \boldsymbol{Z}) = (2\pi\sigma^2)^{-\frac{n}{2}} exp\left[-\frac{1}{2\sigma^2}\left((\boldsymbol{Y}^* - \boldsymbol{Z\beta_z})'(\boldsymbol{Y}^* - \boldsymbol{Z\beta_z}) + k\boldsymbol{\beta}'\boldsymbol{\beta}\right)\right]$$

Dalam penentuan distribusi prior pada penelitian ini menggunakan *conjugate* prior dengan pendekatan *pseudo* prior. *Conjugate* prior menggunakan distribusi eksponensial dalam satu keluarga, sedangkan *pseudo* prior memanfaatkan hasil estimasi sebelumnya yang menggunakan MLE sehingga dapat dimanfaatkan dalam penentuan distribusi prior. Fungsi *likelihood* dan distribusi prior yang sudah terbentuk maka selanjutnya digunakan untuk menentukan distribusi posterior.

2. Estimasi Parameter *Ridge* Bayesian

Apabila kondisi konvergen sudah tercapai maka sampel yang dihasilkan sesuai dengan distribusi target yaitu distribusi posterior, sehingga menunjukkan hasil estimasi yang dihasilkan dari metode Bayesian tersebut sudah mendekati hasil estimasi dengan menggunakan regresi *ridge*. Proses estimasi dilakukan simulasi sampel sebanyak 90000 kali yang memanfaatkan metode *Gibbs Sampling*. Hasil estimasi dengan metode Bayesian pada regresi *ridge* dapat dilihat pada Tabel 4.12.:

Tabel 4.12. Hasil Estimasi Metode Bayesian Data 2

j	**Nilai Estimasi**	***Credible Interval***	
		2.5%	**97.5%**
1	-0.6110	-0.7120	-0.5070
2	0.2520	0.1430	0.3630
3	0.0810	-0.0130	0.1750
4	0.1910	0.0810	0.3030

Hasil estimasi dengan metode Bayesian maka diperoleh bahwa terdapat satu variabel yang tidak berpengaruh, karena nilai *Credible Interval* (CI) menunjukkan nilai yang melewati nol. Pengujian signifikansi pada hasil estimasi dengan metode Bayesian menunjukkan hasil yang sama dengan MLE, hal ini menunjukkan bahwa metode Bayesian sesuai jika digunakan untuk mengestimasi model regresi *ridge.*

3. Pemeriksaan Konvergensi

Pemeriksaan konvergensi pada hasil estimasi dengan metode Bayesian dapat dilakukan dengan tiga cara yaitu dengan memeriksa nilai *MC Error*, ACF dan *Trace Plot*.

Tabel 4.13. Nilai *MC Error* Data 2

j	**sd**	**1%*sd**	***MC Error***
1	0.0515	5.1538 x10^{-4}	2.8282x 10^{-4}
2	0.0565	6.6479 x10^{-4}	3.0280 x 10^{-4}
3	0.0478	4.7859 x10^{-4}	2.4720 x 10^{-4}
4	0.0572	5.7165 x10^{-4}	3.3505 x 10^{-4}

Nilai *MC Error* dikatakan konvergen apabila kurang dari 1% nilai *standard deviasi*, pada Tabel 4.13. menunjukkan bahwa nilai *MC Error* sudah sesuai sehingga dapat dikatakan hasil estimasi sudah konvergen. *Trace* plot menunjukkan sudah tidak membentuk suatu pola tertetu sesuai dengan Lampiran 9 sehingga dapat dikatakan konvergen. Plot ACF dapat dilihat pada Lampiran 10. menunjukkan bahwa *lag* pertama mendekati nilai satu dan *lag* selanjutnya bernilai mendekati nol berarti sudah dapat dikatakan konvergen. Hasil estimasi sudah menunjukkan kondisi yang konvergen sehingga tidak perlu dilakukan penambahan banyaknya iterasi.

4. Transformasi Balik Model Regresi *Ridge* Bayesian

Sama seperti metode regresi *ridge* maka pada metode Bayesian juga perlu dilakukan transformasi balik. Model regresi yang diperoleh adalah sebagai berikut:

$$\hat{Y} = 1.0688 - 0.0138\mathrm{X}_1 + 0.0087\mathrm{X}_2 + 0.1850\mathrm{X}_3 + 0.0438\mathrm{X}_4 \quad (4.6)$$

Model di atas menunjukkan hasil niai estimasi yang tidak jauh berbeda dengan hasil estimasi menggunakan regresi *ridge* pada model estimasi yang dihasilkan persamaan (4.5).

4.2.4. Interpretasi Model dan Pembahasan

Dari hasil esimasi menggunakan regresi *ridge* dan bayesian maka kedua metode menunjukkan nilai estimasi yang tidak jauh berbeda, hasil estimasi dengan kedua metode dapat dilihat pada tabel 4.14.

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

Tabel 4.14. Hasil Estimasi Parameter Data 2

j	**Regresi *Ridge***	**Bayesian**
Intersep	1.0258	1.0688
1	-0.0130	-0.0138
2	0.0086	0.0087
3	0.1843	0.1850
4	0.0446	0.0438

1. Regresi *Ridge*

Sesuai model yang dihasilkan dari regresi *ridge* dengan MLE pada persamaan (4.5) dapat dikatakan bahwa setiap peningkatan 1 km^2 luas wilayah menurut Kecamatan di Kabupaten Tulungagung maka akan menurunkan kepadatan penduduk sebesar 13 jiwa/ km^2 di mana faktor lain dianggap konstan. Setiap kenaikan 1000 jiwa jumlah penduduk maka akan meningkatkan kepadatan penduduk sebesar 9 jiwa/km^2 di mana faktor lain dianggap konstan. Setiap peningkatan 10% laju pertumbuhan penduduk maka akan meningkatkan kepadatan penduduk sekitar 2 jiwa/km^2 di mana faktor lain dianggap konstan. Selanjutnya setiap peningkatan 100 jiwa bayi yang lahir maka akan menyebabkan kepadatan penduduk meningkat sebesar 5 jiwa/km^2.

Pada penelitian ini jumlah penduduk dan jumlah bayi lahir memberikan pengaruh positif terhadap kepadatan penduduk, sedangkan luas wilayah memberikan pengaruh negatif terhadap kepadatan penduduk. Pengaruh negatif dari luas wilayah ini disebabkan oleh luas wilayah yang tidak meningkat seperti grafik pada Lampiran 12, sehingga wilayah tersebut tidak terdapat cukup sumber daya untuk kemajuan penduduknya, misal tidak terdapat perguruan tinggi yang bagus dan lapangan pekerjaan yang tidak mencukupi untuk semua penduduk. Kondisi seperti ini menyebabkan sebagian penduduk pergi ke daerah lain untuk memcari lapangan pekerjaan maupun sekedar meneruskan pendidikan. Beberapa kecamatan tersebut meliputi Boyolangu, Tulungagung dan Kedungwaru. Laju pertumbuhan penduduk tidak memberikan pengaruh secara nyata terhadap kepadatan penduduk dikarenakan pemerintah kabupaten Tulungagung menghendaki untuk menekan laju pertumbuhan penduduk, sedangkan pada

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

beberapa kecamatan dari tahun ke tahun malah menunjukkan laju pertumbuhan penduduk yang meningkat seperti pada Lampiran 13.

2. Regresi *Ridge* Bayesian

Hasil estimasi dengan metode Bayesian pada persamaan (4.6) maka diperoleh hasil tidak jauh berbeda dengan model pada regresi *ridge*. Pada model regresi *ridge* dapat dikatakan bahwa setiap peningkatan 100 km^2 luas wilayah menurut Kecamatan di Kabupaten Tulungagung maka akan menurunkan kepadatan penduduk sebesar 14 jiwa/km^2 di mana faktor lain dianggap konstan, penurunan kepadatan penduduk ini disebabkan karena ketika luas wilayah semakin meningkat namun tidak tersedia lapangan pekerjaan yang memadai untuk penduduknya sehingga menyebabkan penduduk berpindah ke wilayah lain untuk mencari lapangan pekerjaan. Beberapa kecamatan tersebut meliputi kecamatan Boyolangu, Tulungagung dan Kedungwaru. Setiap kenaikan 1000 jiwa banyak penduduk maka akan meningkatkan kepadatan penduduk sebesar 9 jiwa/km^2 di mana faktor lain dianggap konstan. Setiap peningkatan 10% laju pertumbuhan penduduk maka akan meningkatkan kepadatan penduduk sekitar 2 jiwa/km^2 di mana faktor lain dianggap konstan. Selanjutnya setiap peningkatan 100 jiwa bayi yang lahir maka akan menyebabkan kepadatan penduduk meningkat sebesar 4 jiwa/km^2.

Pada penelitian ini laju pertumbuhan penduduk tidak berpengaruh terhadap kepadatan penduduk, hal ini dikarenakan pemerintah Tulungagung mencanangkan program Keluarga Berencana (KB) yang diharapkan mampu menghambat tingkat laju pertumbuhan penduduk. Namun pada kenyataannya laju pertumbuhan penduduk pada tahun 2016 meningkat dari tahun sebelumnya.

4.3. Kriteria Kebaikan Model

Setelah diperoleh hasil estimasi parameter dan hasil pengujian signifikansi parameter maka perlu dilakukan pengujian kebaikan model, di mana dengan beberapa kriteria model yang digunakan dapat menunjukkan hasil yang terbaik dari model yang terbentuk.

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

Tabel 4.15. Nilai Kriteria Kebaikan Model

	Data 1		Data 2	
	MLE	Bayesian	MLE	Bayesian
R^2_{adj}	0.4825	0.7060	0.8746	0.9370
MSE	0.0135	0.0077	0.0078	0.0034
BIC	-148.1597	-168.4788	-89.3977	-99.9784
AIC	-157.9852	-178.3043	-93.1754	-103.7561

Pada penelitian ini menghasilkan dua model estimasi yaitu model regresi *ridge* dengan MLE dan metode Bayesian. Hasil R^2_{adj} menunjukkan pada metode Bayesian menghasilkan nilai yang lebih tinggi. Ketika nilai R^2_{adj} tinggi maka akan menyebabkan nilai MSE cukup rendah, nilai MSE yang cukup rendah mengindikasikan keragaman dalam model kecil sehingga hasil estimasi yang dihasilkan sudah cukup baik. Dalam penelitian ini nilai MSE metode Bayesian lebih kecil daripada MLE. Kemudian hasil nilai BIC dan AIC menunjukkan tingkat ketepatan metode estimasi dalam membentuk model, di mana nilai BIC dan AIC mempertimbangkan fungsi *likelihood* suatu distribusi data sampel. Diperoleh nilai BIC dan AIC pada metode Bayesian lebih kecil daripada metode MLE, sehingga dapat disimpulkan bahwa metode Bayesian cukup baik dalam membentuk model regresi *ridge*.

Dari keempat kriteria kebaikan model yang digunakan, maka dapat diketahui bahwa metode Bayesian lebih baik jika digunakan untuk estimasi parameter dalam regresi *ridge*.

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

BAB V
KESIMPULAN DAN SARAN

5.1. Kesimpulan

1. Model regresi *ridge* dengan estimasi parameter MLE dan Bayesian adalah sebagai berikut:
   a. Data 1 yaitu Angka Harapan Sekolah pada Kabupaten/ Kota di Jawa Timur tahun 2015 sebagai berikut:
   $$\hat{Y} = 6.5642 + 0.0567X_1 + 0.0134X_2 + 0.0265X_3 + 0.0035X_4 + 0.0689X_5 + 0.0239X_6$$
   $$\hat{Y} = 6.1670 + 0.0670X_1 + 0.0150X_2 + 0.0300X_3 + 0.0010X_4 + 0.0820X_5 + 0.0260X_6$$
   b. Data 2 yaitu Kepadatan Penduduk menurut Kecamatan di Kabupaten Tulungagung tahun 2016 sebagai berikut:
   $$\hat{Y} = 1.0258 - 0.0130X_1 + 0.0086X_2 + 0.1843X_3 + 0.0446X_4$$
   $$\hat{Y} = 1.0688 - 0.0138X_1 + 0.0087X_2 + 0.1850X_3 + 0.0440X_4$$
2. Pada penelitian ini diketahui bahwa metode Bayesian lebih baik dalam mengestimasi model regresi *ridge*, karena menghasilkan nilai R^2_{adj} yang lebih besar dan menurunkan nilai MSE, BIC dan AIC.

5.2. Saran

Dari hasil analisis yang telah dilakukan dalam penelitian ini maka saran yang dapat diberikan adalah:

1. Memilih variabel prediktor yang lebih sesuai dalam menjelaskan variabel respon, karena pada penelitian ini terdapat variabel respon yang berasal dari proses perhitungan variabel prediktor yang terlibat dalam model (studi kasus pada Data Kepadatan Penduduk menurut Kecamatan di Kabupaten Tulungagung).
2. Dalam proses simulasi pada metode Bayesian memerlukan kondisi *software R* yang stabil agar pemanggilan *code* saat ini tidak bercampur dengan hasil pemanggilan *code* yang sebelumnya.

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

~ halaman ini sengaja dikosongkan ~

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

DAFTAR PUSTAKA

Abidin, F. P. 2015. *Penanganan Multikolinieritas Dengan Bayesian Ridge Regression Pada Kasus Kemiskinan Kabupaten/Kota Di Jawa Timur*. Skripsi pada Fakultas MIPA Universitas Brawijaya: tidak diterbitkan.

Box, G. E. P and Tiao, G.C. 1973. *Inference In Statistical Analysis*. Addison-Wesley Publishing Company: Inc Philippines.

Carlin, B. P and Louis T. A. 2009. *Bayesian Methods for Data Analysis*. Third Edition. Taylor & Francis Group, LLC: United States of America.

Chatterjee, S., and Hadi Ali.S. 2012. *Regression Analysis by Example Fifth Edition*. John Wiley & Sons, Inc: Hoboken, New Jersey.

Draper, N.R dan H. Smith. 1992. *Applied Regression Analysis, Second Edition*. Terjemahan Bambang Sumantri. John Willey and Sons: New York.

El-Dereni, M., N.I. Rashwan. 2011. *Solving Multicollinearity Problem Using Ridge Regression Models*. Int. J. Contemp. Math. Sciences, Vol.6. No 12, 585 – 600.

Effrihan. *Simulasi Analisis Multikolinieritas pada Regresi Linier Berganda Menggunakan Regresi Ridge Bayesian*. Skripsi pada Fakultas MIPA Universitas Brawijaya: tidak diterbitkan.

Gujarati, Damodar N., dan Dawn C. Porter. 2010. *Basic Econometrics*.5th *Edition*. McGraw-Hill, Salemba Empat: Jakarta.

Greenberg, E. 2008. *Introduction to Bayesian Econometrics*. Cambridge University Press: New York.

Hestie T., James G., Witten D., Tibshirani R. 2008. *The Elements of Statistical Learning-Data Mining, Inference, and Prediction.* Second Edition. Springer: New York.

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

Hestie T., James G., Witten D., Tibshirani R. 2013. *An Introduction to Statistical Learning with Applications in R*. Springer New: York.

Hoerl, A.E., Kennard, R.W. 1970. *Ridge Regression: Biased Estimation for Nonorthogonal Problems*. Journal The American Statstician Volume 12, No.1.

Kutner, Michael H., Nachtsheim C.J., Neter J., and Li W. 2004. *Applied Linear Regression Models Fourth Edition*. New York:Mc.Graw-Hill.

Ntzoufras, I. 2009. *Bayesian Modellig Using WinBUGS*. John Wiley & Sons: Inc, USA.

Pramoedyo, Henny. 2013. *Statistika Inferensia Terapan*. Malang: Danar Wijaya.

Pereira, F. 1999. *Pratical Modern Bayesian Statistics In Actuarial Science*. General Insurance Convention.

Walpole, Ronald E. 1995. *Pengantar Statistik Edisi 3*. Alih Bahasa: Bambang Sumantri. Jakarta: Gramedia Pustaka Utama.

Wathen, S.A., Lind, D. A., Marchal W.G. 2014. *Statistical Teqniques in Bussines & Economics 15thEdition*. Terjemahan Romi Bhakti Hartarto. McGraw-Hill Education, Salemba Empat: Jakarta.

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA